EFFECTS OF ENVIRONMENTAL ENRICHMENT AND BEHAVIOURAL STYLE ON STRESS RESPONSES IN SINGLY HOUSED SHELTER CATS (*FELIS CATUS*)

BY

JACKLYN JAYE ELLIS

A Thesis
Submitted to the Graduate Faculty
in Partial Fulfillment of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

Department of Health Management
Faculty of Veterinary Medicine
University of Prince Edward Island

CONDITIONS FOR THE USE OF THE THESIS

PERMISSION TO USE POSTGRADUATE THESIS

Title of Thesis: “Effects of environmental enrichment and behavioural style on stress responses in singly housed shelter cats (*Felis catus*)”

Name of Author: Jacklyn Jaye Ellis

Department: Health Management

Degree: Doctor of Philosophy Year: 2013

In presenting this thesis in partial fulfillment of the requirements for a postgraduate degree from the University of Prince Edward Island, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the professor or professors who supervised my thesis work, or, in their absence, by the Chair of the Department or the Dean of the Faculty in which my thesis work was done. It is understood any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Prince Edward Island in any scholarly use which may be made of any material in my thesis.

Signature:

Address: Department of Health Management
Atlantic Veterinary College
University of Prince Edward Island
550 University Avenue
Charlottetown, PE, C1A 4P3
Canada

Date:

University of Prince Edward Island

Faculty of Veterinary Medicine

Charlottetown

CERTIFICATION OF THESIS WORK

We, the undersigned, certify that **Jacklyn Jaye Ellis**, a candidate for the degree of Doctor of Philosophy, has presented her thesis with the following title "Effects of environmental enrichment and behavioural style on stress responses in singly housed shelter cats (*Felis catus*)" and that the thesis is acceptable in form and content, and that a satisfactory knowledge of the field covered by the thesis was demonstrated by the candidate through an oral examination held on December 16th, 2013.

Examiners' Names	Examiners' Signatures
Dr. Georgia Mason (External)	___________________________
Dr. Dan Hurnik (Chair)	___________________________
Dr. J. Trenton McClure	___________________________
Dr. Mary McNiven	___________________________
Dr. Cathy Ryan	___________________________

Date:

ABSTRACT

Shelters are potentially stressful environments for singly housed cats. The behaviour and faecal glucocorticoid metabolite (FGM) responses were examined in six shelter cats caged singularly for 30 days. Raised faecal glucocorticoid metabolite (FGM) concentrations and Cat-Stress-Scores (CSS) together with increased grooming and reduced eating indicated a stress response that tended to decrease with time. Environmental enrichment (EE) of the cages may help reduce this stress. However, the efficacy of different types of EE may be reliant on whether a cat is bold or shy, and may be related to the EE preferences expressed. A test was developed to discriminate between bold and shy cats. Nine cats were tested for latency to emerge from a carrier, and percentage of time spent in a carrier during a 5-min test in an empty arena. These measures were found to be significantly different across individuals and not significantly different across time. Seventeen cats were then subjected to a similar test, with similar results. Combining these results, it was determined that latency to emerge from carrier with a cut-point of 10 s was the most appropriate test for discriminating between bold and shy cats in a shelter setting. It correctly classified a high percentage of cats overall, was quick and easy to administer, and best suited to correctly identify shy individuals, which were less represented in this population and arguably could derive greater benefit from identification and extra attention. A final behavioural style study subjected 84 cats to this test, which showed acceptable cross-context consistency and correctly classified a high percentage of cats. Cats' visit frequency to and time allocation with different types of EE were assessed using a plus-shaped choice chamber, in which each external compartment contained a different type of EE. Entrance to each chamber was gained through a cat-flap door monitored remotely using a HOBO data-logger. Usage of the shelf compartment was not significantly different from any other compartment. A significantly higher percentage of time was allocated to the compartment containing the hiding box compared to

the toy compartment and to the empty control compartment, suggesting a hiding box may be a valued resource. A final experiment housed 72 bold or shy cats singly in cages enriched with either a hiding box or a shelf, or in an empty control cage. The amount of food eaten and the percentage of time spent eating increased across time, and percentage of time spent grooming decreased over time. Cats in the hiding box group had significantly lower FGM and higher food intake than cats in the control group, indicating that the type of EE used most in the choice test reduced stress. Shy cats used the hiding box significantly more than bold cats, and had a significantly higher CSS until day 5. There was no evidence that bold and shy cats benefited from different types of EE. The results indicated that the stress of caging is experienced more intensely by shy cats, but it can be reduced by the inclusion of a hiding opportunity in cats expressing either mode of this behavioural style.

ACKNOWLEDGEMENTS

First and foremost I would like to thank my parents Lynda Mayor and Gord Ellis for their love, support, and unending patience throughout not only this degree, but through life in general. Special thanks also go out to supervisory committee, especially my supervisor Dr Michael Cockram and my statistician Dr Henrik Stryhn, but also Dr J Trenton McClure, Dr Cathy Ryan, and Dr Jonathan Spears for the many hours of guidance and encouragement they provided. This project would not have been possible without the funding provided by the Sir James Dunn Animal Welfare Centre and Nutrience pet food. I had assistance working with and transporting the 143 cats involved in this thesis from Vasiliki Protopapadaki, Dave McIver, Dan Hines, Michael Cockram, Jonathan Spears, and particularly from the Prince Edward Island Humane Society (especially Beckie MacLean and Lesa Donnelly) and the Atlantic Veterinary College animal care team (especially Chris McQuaid and Maciej Zawadzki), for which I am very grateful. Thanks also go to Vasiliki Protopapadaki and Samuel Howard for their assistance in watching the mountains of video footage. I would also like to thank Alexander MacNeill, Dan Hines, Matt Hines, Erin McCauley, and Buffy Spoelstra for their help in the construction and design of the plus-shaped choice chambers used in my choice test. Laboratory analysis of the faecal glucocorticoid metabolite concentrations would not have been possible without the advice and assistance of Dr Rupert Palme, Balbir Singh Josen, Dr Mary McNiven, the Atlantic Centre for Comparative Biomedical Research (especially Joy Knight, Kelly Richard, Melissa Perry, and Debra MacDonald), the Marine Natural Products Lab/Kerr Lab (especially Erin McCauley, Jennifer C. Arens, and Dr David Ovary), the Lobster Science Centre (especially Dr Spencer Greenwood, Dr Fraser Clark, Dr Rachel Summerfield, and Dan Hines), and everyone in the Central Services Laboratory. I would like to thank my friends Erin McCauley, Dan Hines, Emily Brown, Dave McIver, Shelby Kerr, Buffy Spoelstra, Alyssa Grunwald, Krista Gill, Megan Bauer, Michael Ciaramella, Steve Bruce, Kady Brown, Logan Otaway, Katherine Duncan, Helen Kennedy, Pat Faller, Andrew Haliday, Michael Rochon, and any other friends or fellow grad students I have not mentioned for their support and amusement over the last four and a half years. Finally, I would like to thank my pets Wyatt (cat #3), Koda, and Indy, for being there for me, no matter what.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACTH=Adrenocorticotropic hormone
ANOVA=Analysis of variance
ANS=Autonomic nervous system
AT=Approach test
BC SPCA=British Columbia Society for the Prevention of Cruelty to Animals
BS=Bold/Shy
CCC=Concordance correlation coefficient
CFHS=Canadian Federation of Humane Societies
CNS=Central nervous system
CSS=Cat-Stress-Score
CTRL=Control group
DD/MM/YY=Day/Month/Year
DF=Degrees of freedom
EAS/EASI=Emotionality, activity and sociability/emotionality, activity, sociability and impulsivity
EE=Environmental enrichment
EIA=Enzyme immunoassay
ET=Emergence test
FGM=Faecal glucocorticoid metabolites
FTP=Feline Temperament Profile
GAS=General adaptation syndrome
GEE=Generalised estimating equation
HPA=Hypothalamic-pituitary-adrenal
ID=Identity
ICC=Intraclass correlation coefficient
IQR=Interquartile range
LED=Light-emitting diode
LHRH=Luteinising hormone releasing hormone
LOS=Length of stay
NA=Not applicable or available
NP=Not possible
OR=Orienting response
PCA=Principal component analysis
PVC=Polyvinyl chloride
SE=Standard error
SNS=Sympathetic nervous system
SPCA=Society for the Prevention of Cruelty to Animals
SD=Standard deviation
UK=United Kingdom
UPEI=University of Prince Edward Island
VOI=Variables of interest

1 INTRODUCTION

1.1 Cats

1.1.1 Cat populations in Canada

The Canadian Animal Health Institute estimated that in 2012 the pet cat (*Felis catus*, henceforth simply referred to by its common name) population in Canada was approximately 7.9 million, and that 37% of Canadian households owned at least one cat. For comparison, the same study estimated that the pet dog population was approximately 6.6 million, and that 36% of Canadian households owned at least one dog (Canadian Animal Health Institute, 2012).

Unfortunately it is difficult to provide accurate Canadian statistics on the population of unowned cats (i.e. stray or feral), however the city of Winnipeg has reported a population of 50 000 (Skerrett, 2011), and estimates in the city of Toronto range from 20 000-100 000 (The Toronto Humane Society, 2011). It is possible to provide relatively reliable statistics on shelter populations. The Canadian Federation of Humane Societies (CFHS) has published a range of statistics online for the years of 1993-2010 on their feline residents – including admission, adoption, and euthanasia rates – (Canadian Federation of Humane Societies, 2010). These statistics were compiled from surveying all Canadian Humane Societies and Societies for the Prevention of Cruelty to Animals (SPCAs), but not municipally-run animal shelters or pounds. While the exact number of respondents varied from year to year, the response rate was approximately 60% in each year. Figure 1.1 shows the total number of cat admissions reported by responding organisations from 1993-2010. For comparison, the total number of dog admissions has also been included.

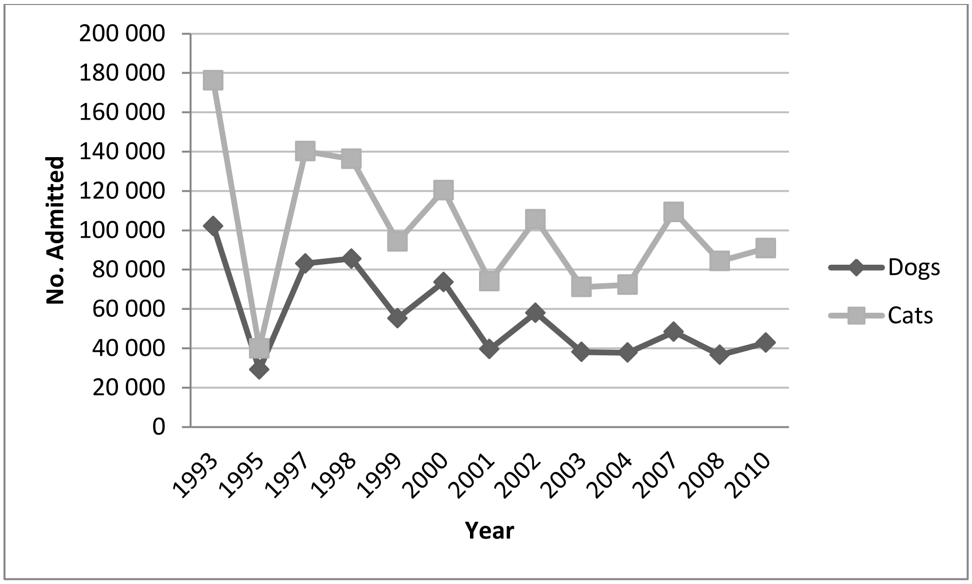

Figure 1.1 Canadian Humane Society reported admission statistics 1993-2010

1.1.2 Cats in shelters

Cats make up a much larger percentage of the population in shelters than do dogs. These cats may be brought in as feral, strays, or surrendered by their owners. The most common reasons why owners relinquish cats include: unwanted litters of kittens, change in owner circumstances, and behavioural problems exhibited by the cat (summarised in Rochlitz, 2000). Figure 1.2 shows final disposition trends for cats from 1993-2010. While adoption rates appear to be on the rise and euthanasia rates appear to be on the decline, euthanasia is still the number one category for disposition of cats both overall and in most years individually. Unfortunately, the statistics provided by the CFHS only broke down specific reasons for cat euthanasia in 2008 and 2010, and even these statistics were difficult to interpret. Categories of reasons included were 'physically and behaviourally unhealthy', 'physically and behaviourally healthy', and 'uncategorised'. The absence of discrimination between physical and behavioural health is surprising, due to the drastic difference in remedial strategies. Wenstrup and Dowidchuk (1999) surveyed 186 American shelters and animal control agencies and reported the number one reason a specific cat would be euthanized was behaviour, while the number one reason a specific dog would be euthanized was health. Interestingly, Weiss (2012) conducted a survey of pet adopters from five American shelters and animal control agencies about the reasons they selected their particular pet. They found that while the number one reason specific cats were selected was behaviour, the number one reason specific dogs were selected was appearance. As the number one reason cats are chosen for euthanasia is behaviour (presumably negative), and the number one reason cats are chosen to be adopted is behaviour (presumably positive), a successful program to reduce a cat's chance of being euthanized, and increase its chances of adoption, must first identify reasons for these problem behaviours. Gourkow and Fraser (2006) imply that these problem behaviours may in part be a product of the shelter environment itself. Shelters present a potentially stressful environment for cats, as they may suffer from an insufficiently enriched environment in terms of social deprivation (Broom and Fraser, 2007) and lack of physical complexity (Carlstead et al., 1993), as well as presenting an unfamiliar novel environment (Kessler and Turner, 1997).

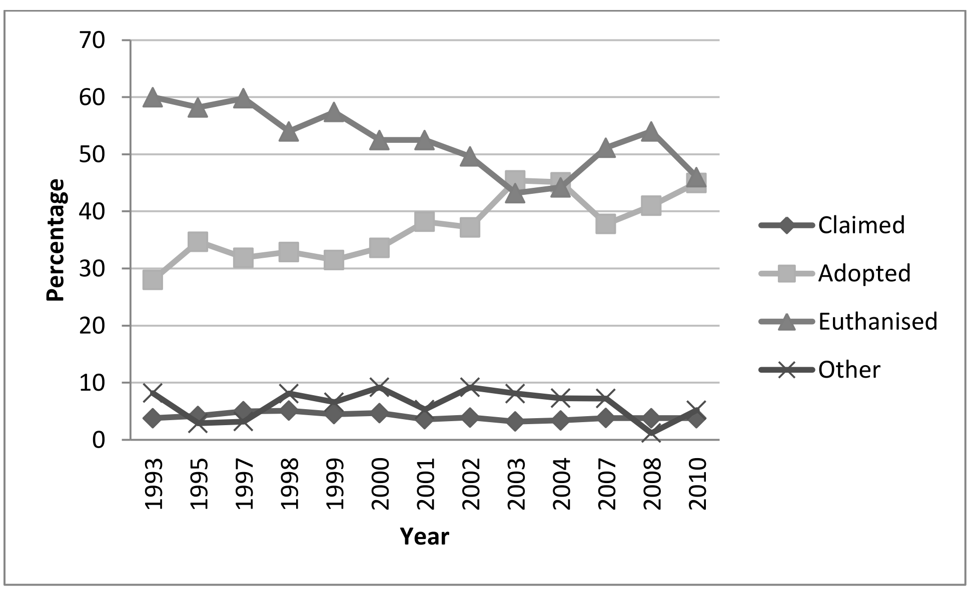

Figure 1.2 Canadian Humane Society cat disposition statistics 1993-2010

1.2 Stress

"Stress" is not as straightforward a concept as many psychological/biological principles, ideas, or disciplines. The term "stress" has been applied both to the force exerted on the organism resulting in an internal strain, and to the strain itself. The common understanding of biological/psychological concept of stress generally refers to the latter (Broom and Johnson, 1993). The major contributors to stress include, but are not limited to: physiological equilibrium, psychological interpretation of situations, and biological responses. For the purposes of this thesis, stress will be defined as a physiological response to perceived threats that have the potential to reduce welfare. These responses can be measured to make inferences about an animal's welfare or affective state. It is important to acknowledge that not all stress is deleterious to the animal. Biological mechanisms have evolved to cope with stress and these biological mechanisms can, in humans be perceived as pleasurable, exhilarating or rewarding (Broom and Johnson, 1993). Researchers use many different methods of studying stress, because it can be expressed by the body through many different response mechanisms.

In order to delve more deeply in to the concept of stress, this section will offer some explanation of the historical development of this concept, as well as the associated biological mechanisms. Furthermore, like the concept of stress, the concept of animal welfare is not straightforward either. As stress is often used as an indicator of negative welfare, the relationship between the two concepts will be discussed. Finally, as methods of assessing indicators of stress can be in themselves stressful, the advantages of assessing physiological indicators of stress non-invasively will be presented.

1.2.1 The history of stress research

The concept of stress as we talk about it today has its origins in the work of Walter B Cannon. Cannon expanded on the views of earlier writers that the most crucial role for survival of living organisms is to keep their internal environment relatively the same, despite changes in the external environment. Cannon subsequently termed the body's ability to maintain its own consistency homeostasis (from the Greek *homoios* similar, and *stasis* position) (Selye, 1976). He did not view homeostasis as something static, just stable. When the homeostasis was threatened, a change would be signalled, and corrective mechanisms would be enacted to avert the threat and restore normality (Cooper and Dewe, 2004). The biggest influence of Cannon's research was his investigation of the sympathetic nervous system (SNS). He contended that in times of challenge the SNS of an animal activates, producing a series of bodily changes that culminate in the "fight-or-flight" response (Archer, 1979). While the biological activity of the SNS will be dealt with more fully in the following section, in essence, the psychoaffective response of the fight-or-flight notion suggests that in response to an immediate threat, an organism responds with fear or anger, which have established associations with instinctive reactions – escape or attack (Cooper and Dewe, 2004). In this way, Cannon introduced the ideas of instinct and emotion into the stress discourse as well. He believed that emotional or social threats could have just as much of a toll on the body as physical threats (Cooper and Dewe, 2004).

Due to his General Adaptation Syndrome (GAS), Hans Selye was probably the most influential figure in the history of stress research. Essentially, this theory states that in response to any noxious agent (or 'stressor') the body produces a non-specific coordinated pattern of defence, organised into three stages. Upon encountering a stressor the body enters stage one: the alarm reaction. This stage represents the call to arms of the body's defence system, and resistance to stress temporarily drops below normal as it absorbs a stressor's initial impact. This stage is controlled by the SNS, and thus is directly analogous to Cannon's fight-or-flight response (Bernstein and Nash, 2008). This initial reaction has the potential to be of such magnitude as to overwhelm the body and result in death. However, in most situations survival occurs, and the alarm reaction is followed by the second stage: the stage of resistance. This stage is characterised by a decrease in production of hormones necessary for growth and reproduction, and an increase in production of hormones necessary for mobilising energy in times of emergency – including the increased production of corticosteroids, such as glucocorticoids (Selye, 1936). During this stage, physiological signs of the alarm reaction are minimised, and the body's resistance to stress increases to above normal levels and stabilises (Bernstein and Nash, 2008). Resistance continues if adaptive energy stores are sufficient to manage the crisis (Selye, 1974). However, if exposure to the stressor continues and the animal is not able to habituate to the stimuli, the animals will enter the third stage: the stage of exhaustion. During this stage the animal loses its acquired resistance and succumbs to the signs expressed in the first stage (Selye, 1936), due to the exhaustion of adaptive energy reserves (Selye, 1974). The name GAS then was a product of the response being *General* to all stimuli, the body's *Adaptive* stimulation of defence in order to cope with the stimuli, and the *Syndrome* of symptoms associated with it (Selye, 1976).

Although Selye's work has not gone unchallenged, his legacy forms the foundations of our modern understanding of stress. The most profound criticisms subsequent researchers had of Selye's work was of the rigid non-specificity of the response to any stressor, and the lack of focus on the animal's perception of the potential stressor. The two most prolific critics of the non-specificity of Selye's GAS were the psychologists John Mason and Richard Lazarus. These researchers stressed the importance of the psychological reaction to stress in the expression of the stress response.

Mason contended that it was not exposure to the stressor itself that resulted in an increase of production and release of glucocorticoids, but the animal's perception of this stressor that was key. He conducted a series of experiments exposing rhesus monkeys to noxious stressors such as extreme heat, moderate exercise, and fasting, but with careful measures taken to avoid the introduction of novelty, sudden changes, or other stimuli thought likely to provoke emotional reaction. Results indicated that a rise in glucocorticoid production was not consistently observed (summarised in Fraser, 2008). This gave strong evidence that the stress response pattern espoused by Selye was too simplistic.

Similar to Mason, Lazarus emphasised the importance of the emotional reaction to the stressor, with focus on the individual's evaluation of a situation. According to Lazarus, the stress response can be hinged on the important appraisal period. Appraisal is the individual's analysis (in either the conscious or cognitive unconscious mind) of whether the demands of an encounter tax or exceed the individual's resources. This process can be broken down further: the primary appraisal concerns discerning whether this event presents any risk or benefit, and the secondary appraisal concerns evaluating strategies to cope with the stressful stimuli (Monat and Lazarus, 1991). This explains how individuals would have dissimilar responses to a similar stressor. While Lazarus conducted his research on human subjects, there is evidence that the appraisal period plays a role in the stress response in animal communities as well. For example, immigrant male baboons are known to kill the infants of lactating mothers, presumably in order to drive the females back into a receptive period. Engh et al. (2006) found that lactating female baboons experienced an increase in glucocorticoid concentrations in response to the arrival of immigrant males, while cycling and pregnant females did not. It is conceivable that upon evaluation of the situation, the lactating mother appraised the immigration as risky and her stress response activated, while the cycling and pregnant females did not appraise the situation as risky, and thus it was not treated as a stressful event.

1.2.2 The biology of stress

Upon encountering a potential stressor, the central nervous system (CNS) integrates and organises the animal's response. Once a new stimulus is introduced, the CNS must first appraise whether to recognise it as a threat to its homeostasis. Perception of the threat – and not the existence of a threat *per se* – initiates the stress response (Moberg, 1985). Stimuli that often trigger this perception include: intense sensory stimulation, pain, startling and/or novel stimulation, and frustration. Some of these rely heavily on complex central nervous processes. If the CNS deems something to be stressful, it must then organise a biological defence in order to maintain its homeostasis. This response can consist of any or all of the following biological responses: behavioural, autonomic, or neuroendocrine (Archer, 1979; Moberg, 1985). There is also some evidence that the immune system may be directly affected by the CNS as a fourth biological response to a stressor, but its effects are largely modulated by the effects of the other responses (Dunn, 1989), and thus will not be treated as a biological response in its own right, but more of a biological consequence of the other responses.

Prolonged exposure to stress can have detrimental effects on the health of the individual, and can leave less energy to be allocated to such processes as reproduction, growth, maintenance, and immune function. In shelter cats specifically, issues commonly associated with stress include feline upper respiratory tract disease (i.e. feline herpesvirus and feline calicivirus) (Edwards et al., 2008; Tanaka et al., 2012), urinary tract disease (i.e. interstitial cystitis and lower urinary tract obstruction), and gastrointestinal disease (i.e. vomiting and decreased appetite) (Stella et al., 2013).

Behavioural response

The behavioural response to stressors is most often an animal's first line of defence, or an initial short term reaction. When a new stimulus indicating possible danger or discomfort is perceived by an animal, the first reaction is often the orienting response (OR). Simply put, this is when the animal turns itself towards the signal, in order to evaluate the signal for its potential as a stressor (Broom and Johnson, 1993). In this sense it may be said that the OR is a behaviour exhibited *before* the CNS determines the stimuli to be a stressor, but if the stimuli is of particularly high intensity, or is painful, the OR may be exhibited simultaneously with a defensive reaction (Archer, 1979). This defensive reaction usually takes the form of withdrawal, immobility, or attack. Exactly which defensive strategy an animal employs depends on a number of factors, including the source of the stress, intensity of the threat, and the assumed consequences. For example, if the stressor is a heat source, withdrawal may be a more effective defence strategy than immobility. If the stressor is a predator that hunts through detecting movement, then immobility may be a more effective strategy than attack. However if the stressor presents a threat to an animal's offspring (and thus its reproductive fitness) attack may present a more effective strategy than withdrawal (Archer, 1979). Behavioural responses to threat represent the most economical response, but are not always practical or possible. Often the behavioural response is to simply remove oneself from the situation, but when an animal's behavioural options are limited – as in captive conditions – this is not always possible (Moberg, 2000).

Autonomic nervous system response

The autonomic nervous system (ANS) has been described as an animal's second line of defence (Moberg, 2000), although some of the chemical changes produced through these pathways help to regulate behavioural responses, for example, the OR (Archer, 1979). Upon encountering a stressor, the sympathetic nervous system (SNS), which is a subsystem of the ANS, produces a series of bodily changes appropriate for muscular exertion – such as increased cardiac output and the adjustment of blood flow towards voluntary muscles – and sources of energy are made available (Archer, 1979). This system provided the basis for Cannon's fight-or-flight response (Moberg, 2000). Many of the physiological responses commonly associated with strong emotions – which can be stressful – are also stimulated by the SNS, such as dry mouth, sweating, pupil dilation, and increased heart rate. Perhaps the most important part of the SNS is the adrenal medulla, the inner part of the adrenal endocrine gland. Stimulation of the adrenal medulla by the SNS produces a class of neurotransmitters/hormones called catecholamines. Common catecholamines include adrenaline (epinephrine) and noradrenaline (norepinephrine), which serve to reinforce the activities of the SNS, in a positive feedback loop. Rises in catecholamines (specifically adrenaline) are associated with novelty, anticipation, unpredictability, and change – situations associated with fear, anxiety, and sometimes anger (Archer, 1979). The physiological effects of these secretions can have drastic effects. Normally very extreme effects are not encountered, as responses of this type tend to be short lived, but in cases of extreme emotional responses to stressors severe SNS activation can occur to the degree that the basic necessary functioning of the body is in danger (Archer, 1979).

Neuroendocrine response

The hypothalamus, a region on the ventral portion of the brain, regulates a wide variety of physiological processes and plays an important role integrating the endocrine system and the ANS. It receives information about environmental conditions from peripheral nerves located throughout the body (via cortical centres), and initiates endocrine responses by secreting a class of hypothalamic neurohormones called releasing and inhibiting hormones (Campbell and Reece, 2005). Each of these releasing and inhibiting hormones has evolved to stimulate or inhibit the excretion of specific hormones by the anterior pituitary (Matteri et al., 2000). The anterior pituitary produces and secretes several different hormones, of which many regulate endocrine glands. Hormones that target endocrine glands are called tropic hormones (Campbell and Reece, 2005).

When an animal encounters a stressor, the CNS signals the hypothalamus to secrete corticotrophin-releasing hormone and vasopressin. These two hypothalamic releasing hormones work independently, and in concert, to stimulate the anterior pituitary to release adrenocorticotropic hormone (ACTH) (Matteri et al., 2000). This tropic hormone in turn targets the adrenal cortex, the outer portion of the adrenal endocrine gland. Upon stimulation, the adrenal cortex synthesises and releases corticosteroids into the bloodstream. There are two main types of corticosteroids: mineralocorticoids and glucocorticoids (Campbell and Reece, 2005). Mineralocorticoids regulate the body's salt and water balance, but glucocorticoids, such as cortisol, mobilise the body's energy sources by regulating the use of carbohydrates, proteins and fats (Oxford University Press, 2000). These resulting intermediaries are then transported to the liver and kidneys, where they are converted to glucose and released into the blood. The increase in circulating glucose provides the extra fuel that can be required for coping with stressors, and is often interpreted as a homeostatic mechanism (Campbell and Reece, 2005). Although changes in circulating concentrations of other hormones – such as prolactin and somatotropin – have proven to also be sensitive indicators of stress, glucocorticoids and the activation of the hypothalamic-pituitary-adrenal (HPA) axis have been the primary neuroendocrine indicators of stress monitored in most studies (Moberg, 2000).

An animal may utilise any or all of these biological responses to a stressor. As Mason and Lazarus pointed out (above), responses to stressors are dependent upon both the type of stressor, and the individual attitudes, expectations, previous experiences, and motives attached to the situation.

1.2.3 Stress as a welfare indicator

Due to the potentially deleterious effects of stress, it is easy to assume that there is always a negative relationship between expression of the various stress responses and the welfare of the animal. However, this relationship is hampered by a number of factors and is far from straightforward. The magnitude and direction of responses often depend on duration, species, age, genetics, experience, as well as context (Veissier and Boissy, 2007).

These factors all contribute to how stressful an animal perceives a situation to be. As stated above, Mason (1971) showed that it was not exposure to noxious stimuli that resulted in activation of the stress response, but perceiving being exposed to noxious stimuli, and the subsequent emotional response. So stress indicators may not necessarily indicate that an animal’s welfare is threatened, but they might indicate that an animal feels that its welfare is threatened. Duncan (1993) argues that it is not how an animal is, but how it feels, that is important to its welfare.

It can be dangerous to simply assume that any one indicator of stress is a direct representation of the welfare of an animal. There is some evidence that whether the animal has control over the situation may influence whether the corticotropic or sympathetic nervous system is triggered. Henry and Stephens (1977) have suggested that when an individual perceives a stimulus as threatening, but within their control to manage, the primary system triggered is the sympathetic adrenal medullary system, but when they sense they have no ability to control the situation, the primary system triggered is the pituitary-adrenal-cortical system. Exactly what is meant by control can vary from being able to avoid, predict, or decide between impending negative stimuli, or being able to encourage, predict, or decide between positive stimuli (Averill, 1973). Expressing displacement behaviours can also be considered an attempt to exhibit control over a situation, even if it functionally accomplished nothing (Dantzer, 1989). This illustrates how measuring signs of the activation of only one system or another may give a very different picture of whether or not the animal is experiencing stress.

Another factor that can shape the stress response is whether the stressor is acute or chronic. For example, while plasma cortisol can be a useful tool for gauging the neuroendocrine response in exposure to acute stressors, there is evidence that if exposure to the stressor is maintained the plasma cortisol concentration will decline after the initial response. This suggests that monitoring plasma cortisol concentration may not be entirely appropriate for assessing chronic stress (Mormède et al., 2007). For example, pigs subjected to long term unpredictable and inescapable electric shocks have similar circulating concentrations of ACTH and cortisol as control pigs (summarised in Mormède et al., 2007). However, while the neuroendocrine stress response showed no difference between treatment and control groups, the behavioural stress response of the two groups were drastically different. This suggests that perhaps behavioural methods can be more appropriate for monitoring chronic stress than neuroendocrine approaches – at least those measuring circulating concentrations in blood and/or plasma samples.

1.2.4 Non-invasive methods of indicating stress

The development of non-invasive techniques has greatly increased the range of hormone studies available to researchers. It is now possible to obtain steroid hormone concentrations from urine, saliva, and faecal samples (Hodges and Heistermann, 2003; Whitten and Brockman, 1998). Collection of these samples is much easier for researchers than sampling blood, but more importantly, such methods eliminate the stress to the animals associated with providing blood samples, and the animals' regular daily activities need not be disrupted. Combining the hormone concentrations obtained using these techniques with behavioural data provides an integrated approach to the study of stress.

Furthermore, unlike more invasive methods, faecal samples reflect cumulative glucocorticoid production over a period of time, and are therefore more representative of general concentrations and are less influenced by specific events likely to register as spikes in production (Palme, 2012). Glucocorticoids are released into the blood stream from seconds to minutes after an encounter with a stressor, meaning measuring plasma concentrations may be best suited to assessing episodic fluctuations. However, circulating glucocorticoids are then metabolised in the liver and "excreted as conjugates via the kidneys into the urine or via bile into the gut" (Touma and Palme, 2005). There is then a lag-time between the time that the glucocorticoids are produced, and the time that their metabolites are excreted in the faecal samples (largely dependent on the species-specific intestinal transit time between the duodenum and the rectum) (Touma and Palme, 2005). As a result of this process, glucocorticoid metabolite concentrations in faecal samples represent the cumulative production, secretion, and elimination of glucocorticoid over a number of hours (Touma and Palme, 2005).

1.2.5 Stress associated with captive conditions

There is a wealth of literature outlining the potential stressors implicit in captive conditions (summarised in Morgan and Tromborg, 2007), and some of the negative effects these stressors can have (summarised in Mason, 2010). In brief, confined animals are exposed to sensory stimuli that can influence their affective state (such as loud or aversive sounds, artificial lighting, odors of prey or predator species, and inappropriate temperatures or substrates), or confinement-specific stressors (such as restricted movement, the inability to retreat, exposure to humans, unpredictable routine, provision of food sources that are not biologically relevant or presented in a non-biologically relevant manner, and reduced or abnormal social groupings). The common theme of both of these types of potential stressors, is the limited control the confined animal has over its environment.

While much of the research in this area is conducted on zoo animals or livestock species, these two classes of stressor have great potential to influence the stress response of cats confined to shelter environments as well. However, these two classes of stressors may be most relevant for animals that are being held in these conditions long-term. While some cats remain in a shelter for long periods of time, others have relatively short stays. Additionally, the initial period in confinement has be described as the period during which cats express the greatest signs of stress, both behaviourally (Hawkins et al., 2004; Kessler and Turner, 1997; McCune, 1992; Smith et al., 1994) and physiologically (Hawkins et al., 2004; Rochlitz et al., 1998). It may then be more relevant to focus on the short term-stressors associated with these shelter conditions: while the above stressors apply, it is likely the novelty of these stressors that impacts the cats most profoundly. The way an individual cat perceives novelty may be mediated by its temperament.

1.3 Temperament

1.3.1 Concept of, and contributions to temperament

It used to be assumed that an individual's behaviour would be dictated by prescriptive categories such as age and sex. Variations from this pattern were simply seen as noise in the data and were considered uninteresting. Slowly, researchers interested in the adaptive significance of alternative modes of behaviour started to recognise that these variations were of themselves worthy of research (Feaver et al., 1986) .

In many studies commenting on the distinctiveness of animals, little distinction is drawn between the terms temperament, personality and individuality. These terms are generally used to describe what distinguishes an animal from other members of its species. An avid observer will come to recognise behavioural patterns exhibited differentially by members of a group, in terms of what behaviours are present, their intensity, and where they are directed. The observer will come to form an impression of the individual distinctiveness of each member of the group in relation to the others based on these patterns. For ease and convenience, these patterns were termed behavioural styles by Feaver et al. (1986) and are often given terms such as 'boldness' or 'friendliness' (Mendl and Harcourt, 2000).

In this way, the temperament of an animal is a mental abstraction on the part of the observer. An easy way to conceptualise this process is to imagine the observation of many discernible behaviours (e.g. fighting, vocalisations, grooming) contributing to the conception of a few behavioural styles (e.g. 'boldness' or 'friendliness'), which contribute to the perception of an individual's temperament (Mendl and Harcourt, 2000).

An animal's behavioural style has great potential to influence its stress response, both by dictating what types of stimuli it perceives as threatening, and by influencing the style of initial response it attempts to mount (e.g. fight or flight). Importantly, as the study of temperament (and the expression of behavioural styles) is the study of the individuality of an animal, it is crucial that if something is to be said to contribute to this individuality, it must maintain some level of stability across time and context. However, it is important to understand that the exhibition of some behaviours may also be age-dependent (such as excitability), and while may remain stable for a short period, could change with extended exposure. Similarly, the exhibition of some behaviours may also be situationally dependent, such as social behaviours, which are likely to be exhibited differently in familiar and unfamiliar social groupings, for example (Stevenson-Hinde, 1983).

Many different genetic and environmental factors have been suggested to contribute to the development of individual differences in behavioural styles and consequently temperament. In the following section some of these factors will be explored, with emphasis on studies specific to domestic cats.

Breed differences

Since pedigree cats represent generations of breeders selecting for particular traits, it is expected that there would be differences between the behaviours of pedigree cats and those of their non-pedigree counterparts. While most studies investigating individuality in cats have focussed on non-pedigree varieties, a few have considered the differences between various pedigree breeds, and between pedigree and non-pedigree domestic cats. Various studies based on questionnaires distributed to either cat show judges (Hart and Hart, 1984) or veterinarians (Fogle, 1991) suggested differences in behavioural styles between different breeds of pedigree cats. For example, Siamese cats were reported to be friendly with strangers, outgoing, and active (Hart and Hart, 1984), Russian Blues were reported to be shy and withdrawn (Hart and Hart, 1984), and Oriental Shorthairs were described as excitable and destructive (Fogle, 1991). Turner (2000a) investigated the owner assessments of various traits of their cats, and also compared these with behavioural observation of the owner's interactions with their cats. This was done for households containing Siamese, Persian, or non-pedigree cats. He found that owners of pedigree cats generally rated their cats higher in traits that could be considered as positive (such as friendliness towards strangers, affection towards owner, and predictability) and lower in traits generally considered negative (such as urine spraying, independence, and aggressiveness) than did the owners of non-pedigree cats. This was corroborated by the behavioural observations, which yielded results indicating pedigree cats spent more time interacting with their owners than non-pedigree (this relationship was stronger for Siamese than Persian), and more time in close proximity to their owners. While the implication so far has been that these behavioural differences were a result of genetic contribution, there is also the possibility that the owners of pedigree cats treat their cats differently than owners of non-pedigree cats (possibly as a result of the increased financial investment), and that this may result in a learned contribution to behavioural differences as well.

Coat colour

A cat's coat colour is dictated by its genetic makeup. Some data have suggested that carrying these genes is linked to more than just phenotypic expression, but to behavioural and distributional differences as well. Todd (1977) found that cats carrying the non-agouti allele were more tolerant of densely populated environments, and thus urban life, than those carrying the agouti allele. Pontier et al. (1995) investigated the difference in distribution of orange cats between urban and rural populations. They found a higher frequency of orange alleles in rural than urban populations. A higher body weight was also associated with this allele. Based on data from another study, stating heavier males were also the most dominant and aggressive (Jones and Horton, 1984), they suggested that in rural polygynous populations a dominant and aggressive reproductive strategy would be advantageous, while in an urban promiscuous mating system this reproductive strategy would not be advantageous, as sperm competition is the primary factor in reproductive success. This would explain the uneven distribution of orange cats between rural and urban settings

Paternity

Domestic cats often have no social contact with their fathers, so while any contribution the mother makes to the offspring's resulting temperament could be due to both a genetic and environmental component, the contribution of the father usually would be strictly genetic. While investigating the behavioural style 'friendliness' (defined as willingness to initiate proximity and/or contact) at two different research sites, Turner et al. (1986) found that kittens ranked as friendly were sired disproportionately between the fathers used in the study. At one of the sites, the father who sired proportionately more friendly kittens could be described as 'friendlier' himself. There was less of a difference between the friendliness of the fathers at the other site, and thus results were less clear. The authors concluded that although it was unlikely that the behavioural style itself was inherited, it was possible that some gene passed on from the father (e.g. growth rate) had an influence on this style through the socialisation process. McCune (1995) similarly found an effect of paternity on the resulting behavioural style of kittens. She found that kittens sired by the friendlier father were friendlier to people, but also were more confident around a novel object. As a result, she reinterpreted the genetic contribution of the friendly father as 'boldness'. Comparably, Reisner et al. (1994) found an effect of paternity on the behaviour of kittens, and when tested, the fathers showed significant differences in the same behaviours. The behaviours showing the most differences were latency to enter test arena, tail position, and response to restraint and venipuncture – all behaviours that could easily be related to temperament.

Handling during sensitive phase

Kittens experience a 'sensitive phase' from 2-8 weeks of age (Karsh and Turner, 1988), during which they form attachments quickly and easily. After this phase, attachments can still be formed or preferences changed, but it requires much more exposure. Evidence suggests that the behavioural style friendliness, especially friendliness towards people, can be directly influenced by handling by humans during this sensitive phase. Karsh (1984) investigated the influence of handling during the sensitive period on friendliness later in life. One of the measures of friendliness she used was the amount of time a cat would tolerate being held. Cats handled during the sensitive phase tolerated being held later in life significantly longer than cats not handled as kittens, cats handled as kittens before the sensitive period, or cats that were handled as kittens slightly after the sensitive period. In McCune's (1995) earlier mentioned study of the effect of paternity on behavioural styles, she also controlled for and investigated the effect of handling during the sensitive phase. Half of McCune's friendly-fathered kittens were handled by humans between 2 and 12 weeks of age, while the other half were not. The same was true for McCune's un-friendly fathered kittens. Similar to the effect of paternity, she found that kittens handled during the sensitive period were friendlier to people, but unlike the paternal effect, no difference was found between the handling groups in terms of confidence around a novel object. Therefore, she concluded that while the genetic paternal effect contributed to 'boldness', the environmental effect of handling contributed to 'friendliness'. In Reisner et al.'s study (1994), kittens were similarly split into handled and un-handled groups, however here no difference was found in tractability later in life between handling groups. This further makes the case for the importance of handling during the sensitive period, as Reisner's handled group was handled between 4 or 5 and 8 weeks of age. While this period is during the sensitive phase, it has been suggested that potential attachment objects must be present at the onset of the sensitive phase to have social behaviour directed towards them (Karsh and Turner, 1988).

Other

A number of other factors have been suggested to contribute to the development of individual differences in behavioural styles, and consequently, temperament. Kittens handled by five people made fewer escape attempts from unfamiliar persons than did kittens handled by one person or no people (Collard, 1967). Being hand raised singly by humans as opposed to by their mother was shown to be related to increased aggression towards both humans and conspecifics (Mellen, 1992). Kittens raised with their mother alone were quicker to emerge into an unfamiliar space than were kittens raised with mother and siblings (Mendl, 1986).

1.3.2 Methods of assessing temperament

As shown by studies presented in the preceding section, there are a range of methods used to assess behavioural styles and temperament. Each method has its advantages and disadvantages, and the methods can be used separately or together to address different types of questions. The following section will outline the four main methods of assessing temperament in feline literature, taken primarily from Mendl and Harcourt (2000) and Manteca and Deag (1993).

Quantitative behavioural observation in a free setting

This method of assessing behavioural styles and temperament essentially consists of observing an animal in an uncontrolled situation, and recording differences between individuals in terms of patterning, frequencies, and durations of behaviours. As the reliability of the data can be tested, this technique has the benefit of providing quantitative data that can be analysed using a range of statistical methods, and is relatively robust. However, while this technique presents the ability to produce reliable data on specific behaviours, it requires substantial inferences on the part of the researcher to relate these specific behaviours to behavioural styles (e.g. boldness, friendliness). Furthermore, the uncontrolled nature of the situation makes it impossible for researchers to explore how these behaviours would be differentially expressed across contexts. Finally, as cross-time consistency must be assessed, several recordings must be performed and the expression of a particular behaviour may be dependent on factors such as age and season. Regardless of these problems, this technique is especially valuable when studying wild populations, when investigating the context of the home environment, or in order to avoid any behavioural abnormalities that may result from an artificial test situation. Cafazzo and Natoli (2009) used this technique to investigate the social function of the tail up posture in cats. While they did not examine individual differences specifically, they did investigate differences in rank, demonstrating the potential to use this technique to investigate other variables that could be used to differentiate between cats.

Quantitative behavioural observation in a structured setting

This method, while similar to the previous method, has a controlled setting. The same quantitative methods of behavioural recording are employed, but the context is dictated by the researcher. An example of this might be how animals behave during exposure to an unfamiliar conspecific. Like the free setting described above, this technique has the benefit of producing a robust, quantitative data set, but it eliminates the lack of cross-context potential as the researcher can alter the context at will. However, this technique still requires substantial inferences on the part of the researcher to relate these specific behaviours to behavioural styles. Additionally, it introduces the problem of potential behavioural abnormalities resulting from an artificial test situation. Finally, although researchers have the ability to alter the stimulants and thus investigate new contexts, the ability to generalise results between different tests or outside of a structured test may be low. Despite these problems this method is widely used due to its ability to help researchers test hypotheses more directly. Mertens and Turner (1989) used this technique to investigate the individual distinctiveness of cats in terms of the behavioural style of 'friendliness towards people'.

Qualitative observer ratings

This method involves two or more observers who have extensive experience with the animals. Each observer rates individuals on specific behaviourally defined behavioural styles. These ratings can then be compared for agreement and reliability. The benefit of this technique is that it has the ability to directly measure the subtle aspects of an individual's behaviour that do not fit well within the structures imposed by conventional recording methods (Lyons, 1989). Additionally, these ratings can be compared with quantitative behavioural observation methods for further validation. The biggest drawback of this method is that it relies on each observer's opinion and is thus subject to personal bias, but the comparison between observer ratings is intended to help assess this problem. Additionally, this technique requires at least two observers who know the animals well, which could reduce the applicability of this method. Feaver et al. (1986) used this method to identify the individual differences of cats in terms of a number of behavioural styles, and the findings were significantly correlated to the results of more quantitative recording methods.

Qualitative owner reports

Essentially this method involves collecting information about the individuality of an animal from the point of view of the owner – presumably the person who has the most intimate knowledge of the animal. The execution of this method can be quite varied, from unstructured written reports, to interviews, to questionnaires. Within human studies of individual characteristics, self-report questionnaires are widely used and are considered quite reliable. In animal studies these reports have the benefit of providing comprehensive descriptions of combinations of individual quirks. However, like observer ratings, personal bias may play a large role in owner description, and unlike observer rating, there are often reports from only one owner, thus making assessment of reliability complicated if not impossible. Additionally, due to the nature of these reports, statistical analysis can also be complicated if not impossible. Meier and Turner (1985) successfully used interviews with cat owners to validate the results of a qualitative experiment, classifying cats into either shy or bold behavioural styles.

1.3.3 Findings from other species

When studying temperament type, it can be advantageous to have an understanding of the findings in a wide range of species. It can be enlightening to see if behavioural styles are universal or species specific. Researchers can learn a great deal by investigating the methods used in other species. With that in mind, included below is a brief description of the leading theories and findings in humans, dogs, and other non-feline animals.

Humans

In humans, temperament is widely defined as the part of the personality reflected by individual differences in reactivity and self-regulation, and is influenced by heredity, maturation, and experience (Rothbart and Derryberry, 1981). In adult humans, the study of personality is far more common than the study of temperament on its own. Cloninger (1994) contends that personality is made up of temperament (unconscious autonomic reactions) and character (conscious, self aware plans), and that temperament is made up of four habit systems or dimensions: 1) harm avoidance, 2) novelty seeking, 3) reward dependence, and 4) persistence. Expression of these dimensions can vary along a continuum, but descriptions of individuals expressing extreme ends of the spectrum are given in Table 1.1. Based on Cloninger's definition of character, it makes sense that only temperament is studied in animals. For similar reasons, temperament (and not character or personality) is studied in infants and very young children. Among the most popular methods used by child psychologists in assessing temperament is the EAS approach, developed by Buss and Plomin (1984; 1986). Essentially, this theory contends that the three broad dimensions of temperament are emotionality, activity, and sociability (hence, EAS). They consider the main criteria of temperament to be inheritance and presence in early life. Originally, their acronym was EASI, also including impulsivity in their list of dimensions, but as the evidence for impulsivity being hereditary was debated, they dropped this dimension in the interest of being scientifically conservative. Table 1.2 gives a description of the definition of the dimensions of EAS.

Table 1.1 Cloninger's Habit systems

Temperament Dimension	Descriptors of High Expression	Descriptors of Low Expression
harm avoidance	pessimistic fearful shy fatigable	optimistic daring outgoing energetic
novelty seeking	exploratory impulsive extravagant irritable	reserved rigid frugal stoical
reward dependence	sentimental open warm sympathetic	critical aloof detached independent
persistence	industrious determined ambitious perfectionist	lazy spoiled underachiever pragmatist

*taken from Cloninger (1994)

Dogs

There has been an abundance of research on temperament in the domestic dog, and it is logical to reference this literature when discussing temperament in cats. As they are both companion animals, temperament research will have similar goals and issues in these species, such as implementing tailored management strategies in shelters, and the nature of the relationship with humans (as compared to wild animals and humans). However, within the dog discourse this research has been conducted by researchers from a range of disciplines with many different goals for the findings (e.g. predicting post-adoption behavioural problems in shelter dogs (Hennessy et al., 2001) and identifying suitable candidates to become guide dogs (Serpell and Hsu, 2001) or police dogs (Slabbert and Odendaal, 1999). As a result, there is a wealth of information, but there is no consensus on the terminology used, making cross-study comparison difficult. And it is clear that some of the goals and issues in the dog discourse would be absent from the cat literature.

Table 1.2 Buss and Plomin's EAS

Temperament Dimension	Descriptors of Extreme High
Emotionality	Defined by a person's tendency to become upset easily and intensely
Activity	Defined by an the level of vigour and tempo with which an individual engages in activity
Sociability	Defined by the tendency to prefer the presence of others to being alone

*adapted from Buss and Plomin (1975; 1984)

Jones and Gosling (2005) conducted a meta-analysis of 51 studies on dog temperament, and used a systematic procedure utilizing expert judges to categorise the concepts discussed in these papers into a standardised lexicon, which has been summarised and presented in Table 1.3. In the years since this review has come out, many more papers have been published on temperament in dogs, but these categories remain reflective of the behavioral styles being described.

Table 1.3 Temperament traits commonly studied in domestic dogs

Temperament Trait	Descriptors of Extreme High
Reactivity	Intense response (positive or negative) to novelty (i.e. activity, approach/avoidance, raised hackles)
Fearfulness	Negative response to novelty (i.e. shaking, avoiding, aggression)
Sociability	Initiating friendly interactions with people or other dogs
Responsiveness to training	Task focused, not easily distracted, 'willing' to work with a person
Aggression	Exhibition of behaviours such as biting, growling, and snapping at people or other dogs
Dominance	'Bullying', food guarding, eating first, refusing to get out of the path of another dog or a person
Activity	Rate of movement

*adapted from Jones and Gosling (2005)

Other species

Across other non-human and non-feline species a range of different behavioural styles (often called temperament traits in this discourse) are identified, but there are five that continually recur: 1) shyness/boldness, 2) exploration/fearfulness, 3) activity, 4) aggressiveness, and 5) sociability. Below these will be defined as presented by Réale et al. (2007). Table 1.4 gives a description of the definition of these dimensions.

Table 1.4 Réale et al.'s five trait system

Temperament Dimension	Description
shyness/boldness	Defined by an individual's reaction to risky situations (i.e. humans or predators), but not new situations (novelty)
exploration/fearfulness	Defined by an individual's reaction to a new situation (that may be risky or safe)
activity	Defined by the individual's general level of activity. As activity in new or potentially risky situations has the potential to be confounded by the previous 2 traits, it is recommended that testing for activity be conducted in the absence of novel or risky stimuli
aggressiveness	Defined by an individual's agonistic reaction towards conspecifics
sociability	Defined by the individual's reaction to the presence or absence of conspecifics, not including aggression

*adapted from Réale et al. (2007)

It is worth noting some striking similarities between Réale et al.'s five trait system and both Coninger's four habits, and Buss and Plomin's EAS. Table 1.5 compares the similar categories between these categorical systems; temperament traits listed in the same row are thought to be comparable. Looking over the definitions provided for each of the categories, the similarities are obvious.

Table 1.5 Similarities between 3 systems of categorising temperament traits

Réale et al.'s five trait system	Coninger's four habits	Buss and Plomin's EAS
shyness/boldness	harm avoidance	
exploration/fearfulness	novelty seeking	
activity	persistence	activity
aggressiveness		
sociability	reward dependence	sociability
		emotionality

1.3.4 Narrative of temperament studies in cats

Although much of the literature investigating temperament in the domestic cat has been touched on in the previous sections, it is worthwhile presenting it in more detail and in more-or-less the chronological order the narrative this field has followed.

For decades studies have investigated the effects of early handling on kitten development (Collard, 1967; Karsh, 1984; Wilson et al., 1965), but not until Meier and Turner (1985) was this effect explicitly linked to behavioural styles. Their study was designed to investigate 1) the reactions of cats to a stranger outside of the home, 2) how those reactions change with repeated encounters, and 3) if, based on these behaviours, cats can be classified on a continuum between shy and trusting. To accomplish this, Meier walked a set route twice daily for the summer of 1983, and whenever a cat was encountered she recorded its reaction to: 1) her presence, 2) her attempt to approach it, 3) her attempt to pet it, 4) her retreat for 5 m and subsequent turn to face the cat. These steps were completed in order, and were only conducted if the cat's reaction to the last step was conducive to continuation (i.e. the cat did not flee). When the same cat was encountered on multiple occasions, reactions were still recorded, but encounter number was noted. Where the cat's owner could be identified, the owner was later interviewed for various types of information, including an assessment of their cat's personality. Results showed that 76% of the 33 cats encountered at least three times had consistent reactions: nine could consistently be petted at stage 3, and 16 would consistently flee at stage 1. These reactions represented close to the extremes available to cats. Based on this and other dichotomising evidence, the authors suggested that cats could be classified as one of two cat types: 'shy' or 'trusting'. In owner interviews, 32 out of 35 owners classified their cat's reaction to strangers as the same type that the authors did based on the reactionary data. This study provided evidence supporting a more or less dichotomised distribution of cats in terms of the behavioural style shy/trusting, which was stable over time, and gave some support for the validity of using owner report to assess behavioural styles.

In a follow-up study to this work, Mertens and Turner (1989) attempted to investigate first encounters between cats and strangers in a laboratory setting, in order to examine the importance of the factors potentially influencing these interactions: sex of the cat, individuality of the cat (defined as consistent individual differences among cats, irrespective of the identity of the human partner), age and sex of the human partner, and activity of the person. The authors organised 240 first encounters between 19 cats and 240 people. The encounters were one of two experiments. Experiment A lasted 10 min and began with the introduction of a cat to a test room in which the test person was already seated in a chair in the centre. For the first 5 min the test person was instructed not to acknowledge the cat. For the remaining 5 min the test person was permitted to interact with the cat in whatever manner they preferred. The behaviour of the cat was recorded continuously. Experiment B lasted 5 min only, and the test person was allowed to engage the cat as soon as it was introduced into the room. The behaviour of the test person was recorded in experiment B. For all of the variables measured in both experiments, the identity of the cat was found to be an equally or more significant factor than any of the other factors investigated, most of which were rarely found to be significant, if at all. The authors also described a dichotomy between the temperaments of cats, describing them as either shy or trusting, in their reactions to initial encounters with the human, when either unresponsive or engaging.

Perhaps the most seminal study of temperament in cats was conducted by Feaver et al. (1986). The authors had felt that each of the cats in their colony had a distinct personality, and that these personalities could be of value in developmental studies of alternative modes of behaviour. The intention of their paper was to investigate the validity of using observer rating to assess the individual distinctiveness of the cats in their colony. Fourteen adult female cats in two groups were observed simultaneously by two observers 2.5 h/day, 5 days/week, for 3 months. Each cat was observed as a focal animal 5 min at a time, and their behaviour was recorded at 15 s sampling intervals, in addition to information regarding proximity to other animals. Additionally, at the end of the 3 months both observers rated each cat on 18 adjectives (e.g. curious, excitable, solitary, etc). Inter-observer agreement scores for each adjective were calculated, and adjectives with agreement less than 0.7 were dropped, leaving seven adjectives: active, curious, equable with cats, fearful of people, sociable with people, hostile to people, and tense. Correlations between the mean of the observers' ratings for the seven remaining adjectives for all cats were then calculated. Categories of adjectives were then formed based on the strength of these correlations (i.e. correlations greater than 0.7 or less than -0.7): 1) active and curious, 2) equable with cats, and 3) sociable with people, fearful of people, hostile to people, and tense. After transforming the ratings into Z-scores they were combined in such a way as to give 'personality scores' for each cat in each of the three categories (active, equable, and social). Similar patterns of personality scores emerged, leading the authors to suggest three types of cats: 1) active, aggressive, bossy cats (positive personality scores for alert and sociable, negative for equable), 2) Timid, nervous (negative personality scores for all three categories), and 3) sociable, confident, easy going (positive personality scores for all three categories). Next, the five observer rating adjectives deemed to have equivalent observational counterparts were compared to these quantitative measures via correlation analysis to assess the validity of these ratings. Correlations were strong and significant when ratings were compared to all behavioural observations, and for a large portion of comparisons to proximity data. Due to the authors' focus on measuring reliability and validity, the results were difficult to refute and this study gave strong evidence for both the potential of observer ratings to assess temperament, and the existence of behavioural styles in domestic cats.

Turner et al. (1986) used aspects of the observer rating system developed by Feaver et al. (1986) to investigate the behavioural style of 'friendliness towards people' in adult female domestic cats and their 3-4 month old offspring at the Cambridge cat colony. After having three people familiar with the cats rate each of them for this variable, the friendliness of the kittens was compared with the friendliness of the mothers; no correlation was found. However, the friendliness of the kittens was disproportionately distributed between the two fathers who sired the six litters. The father who sired proportionately more friendly kittens could be described as 'friendlier' himself. Since the kittens had never met the father, his contribution could only be genetic. The experiment was similarly conducted at a Zurich cat colony. Here results were less clear, but the difference in behavioural style of the two fathers was not as markedly different as in Cambridge. This study was highly influential, as it was the first to show the offspring of one male are reliably different from those of another male, and thus the potential for a genetic contribution to behavioural style or temperament.

McCune (1995), working in the same Cambridge cat colony as Turner et al. (1986) designed a study to investigate both the effect of early handling, and the paternal influence, on friendliness towards humans later in life. She separated litters into four treatment groups of kittens: 1) kittens sired by a friendly father and socialised by humans from 2 to 12 weeks of age, 2) kittens sired by a friendly father and not socialised by humans, 3) kittens sired by an unfriendly father and socialised by humans from 2 to 12 weeks of age, and 4) kittens sired by an unfriendly father and not socialised by humans. At 1 year of age these kittens were all tested for their response to people and to novel objects. She found that groups containing cats sired by a friendly father were quicker to approach, touch, and rub a test person, were more vocal, and spent more time within 1 m of them than groups containing cats sired by an unfriendly father. Friendly fathered cats were also quicker to approach, touch and explore a novel object than those sired by unfriendly fathers. Results from socialised cats/unsocialised cats differed in regards to their response to test persons, but no difference was found between groups in their response to novel objects. She concluded that while early handling by people can contribute to the behavioural style 'friendliness towards humans' later in life, the friendliness of the father contributes to the behavioural style 'boldness', as it is reflected in a general response to novel objects – whether those novel objects are people or inanimate objects. McCune (1992) also reported that in the first 25.5 h of caging, cats had two distinct response strategies. The behavioural differences between active and passive responders are summarised in Table 1.6. Cats from friendly fathers were not only more likely to exhibit behaviours associated with being bold, but were also more likely to be active responders. Similarly, cats from unfriendly fathers were not only less likely to exhibit behaviours associated with being bold (and could thus be termed shy), but were also more likely to be passive responders. This active responder/passive responder dichotomy has support from many other species, as it is analogous to the concept of proactive and reactive copers (summarised in Keeling and Jensen, 2009).

Table 1.6 Active vs. passive responders

	Active responders	Passive responders
Behaviours commonly exhibited in response to the first 25.5 h of confinement	• quicker to emerge into cage • more active • more attempts to escape • more vocal • more likely to groom, feed, urinate and defecate	• slower to emerge into cage • more likely to be immobile and silent • inhibited in maintenance behaviours • more likely to have a flattened posture, to be hiding, and to be shaking

*adapted from (McCune, 1992)

Similarly, Reisner et al. (1994) investigated the influence of handling and paternity of kittens on behavioural styles later in life, but with varying results. The intention of this study was to examine and reduce the stress and aggressiveness associated with fear of handling in adult laboratory cats. Thirteen litters were separated into three groups: 1) removed from their mother at 4/5 weeks of age and handled by humans from then until 8 weeks of age, 2) removed from their mother at 4/5 weeks of age and not handled by humans, 3) a control group that was not removed from their mother and was handled as much as a regular colony cat would be. After 8 weeks, all kittens were moved to group pens and testing began. Testing consisted of a 5-min period in which the test person did not engage the cat (in which several behaviours in addition to latency to approach were measured), a second 5-min test period where the test person put the cat on their lap and attempted to entice it to stay, but no active restraint was used (variables measured were latency to leave the test person’s lap and other proximity data), and a test including a brief examination and jugular venipuncture (during which an ordinal score from 1-5 was assigned for level of tractability). All tests were repeated at 12 and 16 weeks of age. As in McCune et al.’s study (1995), results indicated a difference in behavioural style based on paternity, but unlike the previous study, no effect of early handling was noted. One explanation that has been offered (Turner, 2000b) for the lack of relationship found here between early handling and behavioural style later in life is that the authors started handling their kittens too late in the sensitive phase to form strong attachments for humans, and thus, for the handling to have the desired effect.

Later, Durr and Smith (1997) investigated individual differences of the domestic cat in relation to their social structure. Two multi-male multi-female groups of cats with established social structures were subjected to four tests of social dominance once per week, for 6 weeks: 1) as a group, cats were presented with a novel stimuli (order and latency to approach were recorded, as well as duration of interest); 2) as a group, cats were presented with a familiar food item placed next to a novel stimuli (order and latency to approach were recorded, as well as duration of interest); 3) individually, cats were presented with a novel stimuli (latency to approach was recorded, as well as duration of interest); and 4) as a group, cats were presented with an animal of an unfamiliar species in a cage (latency to approach was recorded, as well as duration of interest). Cats were also subjected to one test of object dominance once per week, for 6 weeks: as a group, after 12 h of food deprivation, cats were presented with a highly desirable food item (cats were ranked from 1-5 on degree of interaction, competitiveness, and willingness to share). The social environment of both groups was regularly subjected to disruption through over-crowding in order to assess whether environmental stability is necessary to maintain individual differences in test scores. Significant correlations were found between individual rankings of latency and duration of interest across time in 13/14 of the measures of social dominance. Correlations were consistently less significant during test 3, when cats were tested individually. Significant correlations were also found across tests between the variables latency to approach, behavioural rank, and duration of interest. Thus, this study showed that domestic cats exhibited consistent individual differences over time and across test situations. As most of the tests were in response to novelty, it can be suggested that the behavioural style the authors were investigating was boldness, if ascribing the definition of boldness provided by McCune (1995). This consistency, despite regular disruptions to the social environment, suggests that stability of the social environment is not necessary to maintain the stability of individual differences.

In 1983, Lee et al. developed a testing procedure to assess a cat's individual distinctiveness in the behavioural styles of sociability, aggressiveness, and adaptability to new situations. The test, called the Feline Temperament Profile (FTP) was originally designed to determine the suitability of individual cats for companion animals in nursing homes, primarily by investigating their reaction to unfamiliar people. However, while the potential for this test to contribute to the narrative of temperament study in cats is undeniable, it went relatively ignored in the field until 2003, when Siegford et al. recognised its potential value – due to ease of administration, small time commitment, and absence of the need of observers with extensive knowledge of the animal – and set about to investigate its validity. The test consists of 10 mini-tests including calling the cat, petting the cat, pulling the cat's tail, and startling the cat. In response to each of these mini-tests, researchers assessed the response of the cat using descriptions provided by the authors. In turn, all of the descriptions are classified into the categories 'acceptable', or 'questionable'. In order to test the validity, Siegford et al. compared FTP scores to data from two quantitative behavioural observations in a structured setting: 1) cats reactions to their caretakers in their colony rooms, and 2) reactions to unfamiliar persons in an open field arena. In addition, basal salivary cortisol concentrations were collected for all cats. FTP scores were found to be significantly different among individuals, to group in three distinct clusters, and to be relatively stable over time. They were also found to be valid when compared to the two quantitative behavioural observations. No relationship was found between FTP and basal salivary cortisol concentrations. In 2011, Iki et al. investigated the FTP's ability to predict behavioural and adrenocortical responses of cats to a mild stressor in the form of a 3-min spray bath. They found that while behavioural and adrenocortical responses were correlated to each other, FTP scores were not related to either measure, and therefore were not a good predictor of a cat's response to this specific stressor. These results show that while the FTP may be a valid test of temperament, there is evidence against temperament type – at least as evaluated by the FTP – being related to stress response.

Early in the 2000s, Lowe and Bradshaw conducted experiments on the effects of early handling on adult behavioural styles (2001; 2002). In 2001, they investigated behaviour directly following feeding, and using principal components analysis found four different behavioural styles consistent during all or part of the period from 4 months to 2 years of age: staying indoors, rubbing, investigative, and boldness. They found that early handling had the highest effect on the boldness behavioural style, as kittens receiving the most handling tended to score the highest on the boldness component. In 2002, they tested cats' tolerance of being held by an unfamiliar person by measuring escape attempts, and compared it to the amount of handling they received at an early age. They found that while at 2 months of age the number of escape attempts was highest by kittens that were handled the least, at 4 months of age this trend had reversed. However, they suggested that this was because the escape attempts at 4 months were less likely an attempt to escape the person, and more likely an attempt to initiate play. The authors concluded that the behaviour of a cat when handled by an unfamiliar person is a stable character trait of its own, and that it may be associated with extensive early handling.

Each of these studies investigated temperament in cats, but they used drastically different terminology and methodology. However, despite these differences closer analysis reveals that only a few behavioural styles were investigated; most of which recur across many of the studies. Table 1.7 shows how the different behavioural styles investigated across the studies clump together, and provides encompassing terms for these clumps for further discussion.

Table 1.7 Behavioural styles investigated in studies of feline individuality

Study	Behavioural styles investigated (terms used by authors)			
(Lee et al., 1983)	friendliness towards humans			
(Meier and Turner, 1985)	shy/ trusting (of humans)			
(Mertens and Turner, 1989)	shy/ trusting (of humans)			
(Feaver et al., 1986)	sociable (with people)		alert (active + curious)	equable (with cats)
(Turner et al., 1986)	friendliness towards humans			
(McCune, 1995)	friendliness towards humans	bold		
(Reisner et al., 1994)	friendliness towards humans	bold		
(Durr and Smith, 1997)		boldness		
(Siegford et al., 2003)	friendliness towards humans			
(Lowe and Bradshaw, 2001)	friendliness towards humans			
(Lowe and Bradshaw, 2002)	friendliness towards humans			
(Iki et al., 2011)	friendliness towards humans			
Summary term for style:	**friendliness towards humans**	**shyness/boldness**	**alert**	**equable with cats**

Clearly friendliness towards humans is the most commonly studied behavioural style, which makes sense considering that the cat is a companion animal. The second most commonly studied behavioural style is shyness/boldness. Using McCune's (1995) definition of this behavioural style – general response to novelty irrespective of whether the novelty is human or object – we could consider any of the studies that investigated friendliness towards humans as also investigating shyness/boldness. In fact, Meier and Turner (1985) and Mertens and Turner (1989) even called the behavioural style they were investigating shy/trusting, which has obvious parallels with the behavioural style shyness/boldness. Although Feaver et al.'s (1986) behavioural styles 'alert' and 'equable with cats' were not investigated in any of the other studies, the fact that they were found to be reliable, valid, and uncorrelated may suggest they merit further investigation.

After examination of the different behavioural styles investigated in cats, it can be enlightening to compare this to behavioural styles commonly investigated in humans and other species. Table 1.8 is a modified version of Table 1.5, now including behavioural styles investigated in cats in the comparison. It is worth noting that while studies of individuality in cats use the term behavioural styles, and these other studies use the terms temperament traits or dimensions, these two terms do not differ in functional meaning.

Table 1.8 Similarities between three systems of categorising temperament traits/behavioural styles

Réale et al.'s five trait system	Coninger's four Habits	Buss and Plomin's EAS	Behavioural styles investigated in cats
shyness/boldness	harm avoidance		shyness/boldness, friendliness towards humans
exploration/fearfulness	novelty seeking		alert, shyness/boldness, friendliness towards humans
activity	persistence	activity	alert
aggressiveness			
sociability	reward dependence	sociability	
		emotionality	equable with cats

There are some clear similarities between the temperament traits discussed in these other systems, and the behavioural styles investigated in studies of individuality in cats. It may have seemed intuitive to align friendliness towards humans with sociability; however Réale et al. specifically state that they are referring to intra-species interactions in their definition of sociability. Instead, friendliness towards people is clumped with both shyness/boldness and exploration/fearfulness in Réale et al.'s system, as the former specifically mentions humans, but the latter refers to situations in which the person may be a stranger (and thus, novel). Since McCune (1995) found no difference in expression of 'friendliness towards people' depending on whether the person was familiar or a stranger, the behavioural style 'friendliness towards people' must clump with both of Réale et al.'s traits. The behavioural style shyness/boldness also clumps on two rows. This is because of unfortunate differences in definitions. For example, Réale et al. define shyness/boldness by an individual's reaction to risky situations (i.e. humans or predators), but not new situations (novelty), and define exploration/fearfulness by an individual's reaction to a new situation (that may be risky or safe). However, as McCune (1995) defines shyness/boldness by an individual's reaction to novelty irrespective of if the novelty is human or object (and, presumably, whether it is risky or safe), this clearly encompasses both Réale et al.'s trait of shyness/boldness, and their trait of exploration/fearfulness. While shifting McCune's definition to be in line with the one used in other species might seem simplest, her definition was not based on her subjective impressions alone, but on the results of her experiments. Feaver et al.'s behavioural style alert also clumps on two rows. Since this behavioural style is essentially the combination of active and curious, it makes sense to clump the active part with the activity temperament trait present in both Réale et al.'s five trait system and Buss and Plomin's EAS, while it makes sense to clump the curious part with Réale et al.'s exploration/fearfulness and Cloninger's novelty seeking. Feaver et el.'s equable with cats clumps nicely with Buss and Plomin's emotionality, as both refer to the individual's ease of arousal or tendency to react strongly. While it is clear that there are many strong similarities between the behavioural styles investigated in cats and the temperament traits investigated in humans and other animals, it is also quite evident that the categories do not directly compare.

1.4 Animal welfare and environmental enrichment

1.4.1 Definition of terms

The term 'animal welfare' is difficult to define. A common focus is on the biological functioning of the animal, and suggests the animal's ability to function biologically within its evolutionarily selected limits as evidence for its welfare. Others focus on the animal's ability to cope, advocating for the animal's ability to maintain its homeostasis in response to environmental challenge as the litmus to its welfare. Others still use a mentalistic approach to assessing the welfare of an animal, highlighting the importance of how an animal 'feels' about itself and its environment (Duncan and Fraser, 1997). Each of these definitions has merit. Here, animal welfare will be defined as both the physical and psychological well-being of an animal.

From an animal welfare point of view, the use of animals by various factions of society is acceptable, providing the physical and psychological well-being is maintained (Young, 2003). Unfortunately without guidance, it can be difficult to anticipate what would constitute a threat to an animal's psychological well-being – and to a lesser extent, its physical well-being – so an attempt has been made to create a template for anticipating these needs, in an effort to standardise the animal welfare protocols of various animal keeping organisations. The Five Freedoms were originally proposed by Brambell (1985) in relation to space requirements. They were then developed by the Farm Animal Welfare Council (2009) (http://www.fawc.org.uk/freedoms.htm) into a framework of freedoms that describe an ideal state of welfare rather than standards for acceptable welfare. . They are: 1) freedom from hunger and thirst; 2) freedom from discomfort; 3) freedom from pain, injury, and disease; 4) freedom to express normal behaviour; and 5) freedom from fear and distress. Most reputable animal keeping organisations have adopted these ideas as a good starting point for assessing the welfare of the animals in their care; if an animal in their care can be said to have all or most of these freedoms, then their welfare is generally good.

Sometimes however, particularly in animal research, it becomes necessary to restrict one or more of these freedoms. The acceptability of these restrictions is contentious in society, but within animal welfare research it is generally considered acceptable if the cost to the individual animal is outweighed by the potential benefit of the study. Even then, most organisations require any researcher using animals – particularly those restricting their freedoms – to consider alternative models of inquiry. The standard method of evaluating the appropriate use of animals in research was developed by the Universities Federation of Animal Welfare in the 1950s and is termed the three Rs (Russell and Burch, 1959). They are: 1) replacement – the idea that the use of animals may be replaceable by other models, perhaps computer models or simulations, or the replacement of higher vertebrate species with less sentient animals; 2) reduction – the idea that if animals must be used, careful consideration should be given to how many, and the bare minimum should be employed; and 3) refinement – the idea that modifications can be made to husbandry practices and experimental procedures that may reduce pain, distress, and the restriction of the Five Freedoms. An important component of refinement is the provision of environmental enrichment (Young, 2003), which will be defined and discussed below.

It is important to have some concept of how much an animal's welfare is being compromised, but as the concept itself is somewhat of a mental abstraction, measuring an animal's welfare directly is impossible. Direct observations about the behaviour, physiology, psychology, and neurology represent the most accessible method for indirectly assessing the way an animal is coping with its environment (Broom, 1988).

A combination of methods is likely the most effective strategy to indirectly assess an animal's welfare, because coping methods vary between individuals; the absence of evidence of compromised welfare using one method may not mean that there would not be evidence using another (Broom, 1988).

Similar to the term animal welfare, the term 'environmental enrichment' is difficult to define. This term has been defined variably and used inconsistently in different sources. There is some debate as to whether the term should be applied to the change in the environment itself, should be reserved for any positive outcomes resulting from the change (Newberry, 1995), or should refer only to additions to the environment after the physiological and behavioural needs of the animal are met (Duncan and Olsson, 2001) (this concept will be discussed further in Ch 5 and Ch 9). Here, the term will be defined as any addition to the environment of an animal resulting in a presumed increase in the environment's quality, and a subsequent presumed improvement to the animal's welfare. The resulting 'presumed increase in environmental quality' can then be assessed in terms of its 'improvement to the animal's welfare' by measuring changes in that animal's behavioural, physiological, psychological, and neurologic responses to their situation. This practical definition has been chosen because environmental enrichment is often employed in settings where thorough investigations of the actual benefit are impossible. Given the evidence presented below, it seems reasonable to assert that where evaluation of the actual benefits of environmental enrichment is not possible, making changes to the environmental quality that are only *presumed* to benefit the individuals is still desirable.

Today, environmental enrichment is employed to varying degrees in a range of environments, such as zoos, farms, animal shelters, animal research facilities, and even private homes. Historically however, the introduction of formal environmental enrichment programs is a relatively new advent. Although collections of animals have been kept since antiquity (Young, 2003), historically zoos have not been well known for the treatment – and particularly housing – of their animals. The first examples of environmental enrichment emerged in zoos more to appease the public than to improve the welfare of the animals. In some instances, the modifications made to enclosures only managed to prevent the expression of abnormal behaviours without treating the cause, and thus but made no improvement to the animal's welfare (Young, 2003). Eventually, through the research and writings of groundbreaking psychologists and biologists such as Lorenz, Harlow, Morris, and Hediger (summarised in Shepherdson, 1998), arguments were made for the potential of environmental enrichment to actually benefit the animals in zoological and laboratory settings, and in some instances gave the first evidence to this effect from systematic experiments. Due to the pioneering work of these researchers, the application of environmental enrichment has come a long way. Indeed, providing environmental enrichment for captive primates is not only commonplace, but actually in some countries mandated by legislation (Shepherdson, 1998). Additionally, according to a meta-analysis conducted by de Azevedo et al. (2007), the number of articles published about environmental enrichment in peer-reviewed journals accessed by the database Web of Science experienced exponential growth in the 10 years prior to their paper.

Environmental enrichment has the ability to affect the welfare of an animal through a number of different potential mechanisms. Markowitz published a series of papers (summarised in Markowitz and Aday, 1998) emphasising the potential for environmental enrichment (largely feeding enrichment) to increase the amount of control an animal had over its environment and daily routine. The implication being that these 'empowered' animals experienced greater psychological welfare. In a review by Hughes and Duncan (1988), evidence is presented that animals will perform behaviours in captivity that are not necessary (e.g. hens will perform nest building sequences, even when their previously created nest is still available). The authors contend that animals are biologically motivated to perform these behaviours, and therefore may become frustrated when unable to perform them. Environmental enrichment was suggested as a mechanism to allow for, and even encourage, these behaviours. Wood-Gush and Vestergaard (1989) discuss the importance of exploratory behaviour to an animal, and that the restriction of opportunities to exhibit this behaviour may result in frustration. Such frustration could arise from not being able to perform a highly motivated behaviour, and from boredom since it is not able to keep itself occupied. As captive conditions have already been established as potentially stressful, and activation of the stress response has been shown to hinge upon the perception of a situation as being stressful by the individual, it is possible that by enriching a captive environment appropriately, caretakers may eliminate or reduce the associated stress by influencing the way the individual appraises its circumstances.

1.4.2 Types of environmental enrichment

Environmental enrichment can be broken down into types in several ways based on the way they are conceptualised by different researchers, and specific enrichment items can often be classified into several different categories. Below, an attempt will be made to break down the different types of environmental enrichment – synthesising the classificatory systems presented by a range of authors.

Social enrichment

According to the data presented in the meta-analysis conducted by de Azevedo et al. (2007), social enrichment comprised 15% of environmental enrichment studies published between 1985 and 2004 that focused on a singular type of enrichment. This type of enrichment can be further broken down into the categories of intra-specific and inter-specific (Ellis, 2009). Intra-specific social enrichment refers to social interactions with other individuals of the same species. Alternatively, inter-specific enrichment refers to social interactions with other species: most commonly humans, but on occasion this can refer to interaction with other non-human species. The effect of human interaction with animals has been investigated in many species. This type of interaction is especially valuable when solitary housing is unavoidable (Young, 2003), and can include training and companionship. It has been suggested that human contact is more beneficial to the well-being of shelter dogs than contact with other dogs (Wells, 2004). The effect of interaction with other non-human-species has been studied less often, and generally includes communal housing of compatible species, and companionship (Young, 2003).

Feeding enrichment

de Azevedo et al. (2007) presented data suggesting feeding enrichment represented 13% of the reviewed environmental enrichment studies that focused on one type of enrichment. Feeding enrichment is an alternative to the artificial methods of food presentation traditionally employed in captive environments (i.e. provision of nutritious – but not biologically relevant – food sources in an organised, concentrated presentation). It can consist of a range of deviations from conventional food provision methods, such as live prey, carcasses of relevant prey species, scattering small food items throughout browse materials (thus encouraging foraging), or complex feeding toys/apparatuses.

Sensory enrichment

The data presented in de Azevedo et al. (2007) suggest sensory enrichment represented 16% of the reviewed environmental enrichment studies that focused on one type of enrichment. The senses of non-human animals are often much more keenly developed than our own, and the sensory world represents an opportunity for greatly enriching the environment of an animal. This type of enrichment refers to sensory modifications/additions of the animal's enclosure including olfactory, auditory, visual, gustatory, and textural factors.

Physical enrichment

Unfortunately, it is impossible to report the percentage of environmental enrichment studies reviewed by de Azevedo et al. (2007) on this type of enrichment represents, due to the difference in classificatory systems presented by those authors and here, however the category that contains physical enrichment was the largest category of enrichment presented. Physical environmental enrichment refers to modifications of the animal's enclosure including size, furniture, and use of space.

Enrichment with toys, cognitive opportunities, and other novelties

Again, due to differences in classification, it is impossible to report the percentage of environmental enrichment this type of enrichment represents in the studies reviewed by de Azevedo et al. (2007). In this category, several terms have been lumped together. This is because either the terms are difficult to define, or the description of an item depends more on the academic motivation of the researcher than the actual difference it makes to the animal. For example, any novel object or cognitive puzzle could elicit play, and thus be considered a toy. Examples of this type of enrichment include objects simulating prey, mechanisms that require animals to work for rewards, and objects that are changed with some frequency.

Other

On occasion, other factors have been considered as environmental enrichment. While the description of environmental enrichment itself does not preclude these factors from being included, they do not fall within the traditionally considered categories. An example of this would be regular routine.

Combinations

Commonly, types of environmental enrichment are combined, due to necessity (i.e. feeding toys) or because the experimenter is interested in the general effect of environmental enrichment, and not in the differential effects of the specific type. It has also been suggested (van Praag et al., 2000) that it is the interaction of various types that is the essential to the effects of environmental enrichment, and not a single element hidden in the complexity.

1.4.3 Effect of environmental enrichment

Having delineated the different types of environmental enrichment, it is now possible to evaluate instances of their implementation in terms of their ‘presumed increase in environmental quality’ by looking at the success they had making ‘improvement to the animal’s welfare’. This improvement can be assessed by measuring changes in that animal’s behavioural, physiological, psychological, and neurologic responses to their situation. This has been presented below using examples from various species and captive conditions.

Behavioural function

Common behavioural methods used to make inferences about an animal's welfare include: behavioural observations of the presence of abnormal behaviours or normal behaviours expressed in abnormal frequencies, ordinal rating systems, and behavioural testing.

The most obvious instances of environmental enrichment having a positive influence on the behavioural function of an animal are when animals exhibited extreme negative behaviours, such as injurious or stereotypical behaviours. Using behavioural observations, Martrenchar et al. (2001) were able to show that the addition of straw and metal objects to the environment of turkeys resulted in significantly lower levels of injurious feather pecking in males and females, by redirecting the pecking behaviour. Reed et al. (1993), using a combination of behavioural tests and an ordinal rating system, found that hens that had been enriched with a variety of enrichment types, but notably early socialisation with humans, scored significantly lower 'potential damage risk scores' when exposed to novel or potentially threatening stimuli than similar individuals who had not been enriched. According to a meta-analysis of enrichment studies carried out at zoos conducted by Shyne (2006) in which 54 studies were analysed, 90% showed a reduction in stereotypic behaviour in relation to the baseline. Baker (2004) demonstrated that only 10 min per day of human contact was enough to reduce abnormal behaviour, such as regurgitation and reingestion, in chimpanzees. Using behavioural observations, Wells and Irwin (2008) found that elephants provided with auditory enrichment in the form of classical music spent significantly less time engaged in stereotypic behaviours than they did during control periods. Schipper et al. (2008), using behavioural observations, found that providing kennelled dogs with feeding enrichment toys resulted in a decrease in inactivity, and the appearance of more variable behaviour patterns. These studies illustrate examples of the positive influence each type of environmental enrichment has been shown to have on the behaviours of a range of species in various captive conditions, employing various methods of evaluation.

Physiological function

Although much of the evidence indicating environmental enrichment can improve an animal's welfare is behavioural, hormonal indicators (such as changes in activation of the HPA axis and the SNS) and improved immune function are commonly used physiological methods. Belz et al. (2003) found that rats provided with toys and structural enrichment had lower circulating plasma concentrations of ACTH and corticosterone than control groups, while Naka et al. (2002) found that mice provided with toys and structural enrichment had higher concentrations of noradrenaline in their parieto-temporo-occipital cortex, the cerebellum, and the pons/medulla oblongata than did control groups. Schapiro et al. (2000) investigated the impact of social enrichment on the immune function of rhesus macaques. He found that monkeys housed alone had significantly lower $CD4^+$ to $CD8^+$ ratios than did monkeys housed in pairs or in groups, and that lymphocyte proliferation responses to pathogens were higher in group housed monkeys than in monkeys housed singly or in pairs. The results of this study suggested enhanced immune responses for animals of this species provided with social enrichment.

Psychological function

Psychological methods used to make inferences about an animal's welfare include: behavioural observations and behavioural testing. Markowitz (summarised in Carlstead and Shepherdson,

2000) conducted a number of studies affording control of some aspect of their environment to zoo animals with promising results. One such finding was reduced aggression and stereotypic behaviour in a mandrill given the opportunity to play a computer game with zoo visitors that resulted in a food reward. This increase of control is thought to result in a reduction of frustration. In a review by Hughes and Duncan (1988), evidence is presented of some behaviours (particularly appetitive) that animals are motivated to exhibit even in the absence of the need to achieve the apparent goal. They argue that preventing performance of these behaviours could result in frustration and stress, and thus conversely, allowing or encouraging these behaviours could result in improved psychological welfare. Additionally, Wemelsfelder (1997) has argued that the lack of performing an inhibited, but highly motivated behaviour may result in a void of activity that is filled with boredom.

Neurological function

Common neurological methods used to make inferences about the efficacy of environmental enrichment include: anatomical changes, improved memory and learning, electro physiological changes, and amount of neurotransmitters. Environmental enrichment was found to have a positive influence on brain weight in rats, mice, and gerbils by Rosenzweig and Bennett (1969), and on brain size in rats by Diamond et al. (1966) and Altman et al. (1968). Social enrichment was found to have a positive influence on learning in a maze test and to alleviate deficiencies in learning caused by brain lesions by Pacteau et al. (1989). Larsson et al. (2002) tested the ability of rats to complete two cognitive tasks – a water maze task and open field test – after exposure to a mild and powerful stressor. Thirty-two rats were housed in environments enriched with a variety of objects, while 32 rats were housed in relatively impoverished environments. Enriched rats showed enhanced escape performance in the water maze task, with prior exposure to both a mild and powerful stressor. Enriched rats also exhibited lower levels of activity and fewer defecations than the impoverished group, indicating less emotional reactivity to a novel environment.

Interpreting these findings

The studies reported above represent only instances in which various types of enrichment and methods of assessment have resulted in changes in function. There are also many examples of studies in which no change in behavioural, physiological, or neurological parameters were detected, and indeed, some instances in which these parameters were negatively affected. For example, Moncek et al. (2004) found that rats provided with environmental enrichment had larger adrenal glands and increased adrenocortical function compared to rats housed without environmental enrichment – potentially indicating increased stress. Regardless, the evidence above suggests that there is basis for the assertion that environmental enrichment – when appropriately selected – has the potential to positively influence the welfare of captive animals. However, some of the changes in behaviour, physiology or psychological state have questionable construct validity as measures that necessarily indicate improvement to welfare. It can be dangerous to assume that any change associated with environmental enrichment is necessarily an indicator of improved welfare. For example, above it was reported that Naka et al. (2002) found increased noradrenaline in the brain associated with the provision of environmental enrichment. Would such an increase be interpreted as an improvement of welfare if it was not associated with environmental enrichment, which is presumed to have a positive impact? In order to objectively assess the impact of environmental enrichment on the welfare of animals, it is essential to determine what measures have the construct validity to indicate improved welfare *a priori* in order to avoid making inferences that are invalid and irrelevant.

1.4.4 Evidence from the feline literature

A range of methods have been used to investigate the efficacy of different types of environmental enrichment on the welfare of the domestic cat. Ellis (2009) suggested that opportunity for contact with conspecifics has the potential to have a positive influence on a cat's welfare, provided sufficient space and resources exist to support more than one cat. However, it is worth noting that this effect is dependent upon the previous experiences of the individual cat. Kessler and Turner (1999) found that during their stay at an animal shelter, cats that had not been previously socialised to other cats scored significantly higher on the ordinal Cat-Stress-Score (Kessler and Turner, 1997) when housed with other cats than did cats that had been previously socialised to other cats. Similar results have been found for social enrichment with human contact. The effect of human presence is dependent upon the previous experiences of the individual; differences in the later responses to humans of cats handled and unhandled as kittens are evidence of this (Karsh, 1984; McCune, 1995).

The effect of feeding enrichment on the management of obesity in domestic cats was investigated by Clarke et al. (2005). Cats were subjected to three voluntary exercise periods daily, two involving enrichment with food rewards and one involving interaction with enrichment items not involving food. Half of each cat's food was offered in standard presentation, while the other half of its daily food was offered via the feeding enrichment. Data from daily weighing and body condition scores suggested that obese cats lost a significant amount of weight.

In the domestic cat, structural enrichment in the form of surfaces for claw abrasion (i.e. scratching posts) and enabling hiding behaviours are often recommended (e.g. Rochlitz, 1999). Durman (1991; in Smith et al., 1994) investigated the behavioural changes of communally caged cats during the first month at a shelter for his BSc thesis. He found – among other things - that the proportion of time spent hiding under things dramatically declined after the first four days, and experienced another decline after 15 days. As some studies have suggested that initial placement in a shelter can be the most stressful time for cats (e.g. Kessler and Turner, 1997) perhaps hiding is a way of coping with this stressor. In a study by Kry and Casey (2007), shelter cats provided with a hiding opportunity (in the form of a British Columbia SPCA Hide and Perch box) scored significantly lower Cat-Stress-Scores, were significantly more likely to approach a researcher during an approach test, and slept restfully significantly more often than the control group. There was no significant difference in adoption rates between the 2 groups. Griffith et al. (2000) conducted two experiments involving a feline facial pheromone supposed to have an anxiolytic effect. In the first experiment, 10 cats were exposed to the pheromone and 10 cats to a placebo, and behaviours and food intake were monitored for 90 min. Cats exposed to the pheromone exhibited a significantly higher frequency of lying down, sitting, grooming, interest in food, and eating, and significantly lower frequencies of sleeping. However, no significant difference in mean food intake was observed between the two groups. In the second study, 10 cats were exposed to the pheromone and given access to a cat carrier that they could hide in or sit on, and 10 cats were exposed to the pheromone alone. Food consumption was monitored for 24 h. In this study, the cats in the pheromone and cat carrier group consumed significantly more food than the pheromone only group. The authors suggested that the increase in grooming, interest in food, and food intake are evidence of the anxiolytic effect of the pheromone, and that the further increase in eating accompanied by the provision of the cat carrier may serve to enhance the effect of exposure to this acute ‘anti-stressor’.

Both auditory and olfactory enrichment has been investigated in the domestic cat by Ellis and Wells (2008; 2010). Visual stimulation in the form of television images – notably combining prey items and linear movement – elicited interest and altered behaviour of shelter cats, and the authors advocated future study of its potential for successful environmental enrichment in cats (Ellis and Wells, 2008). Olfactory stimulation – notably prey scents and catnip – elicited significantly lower levels of inactivity and catnip additionally elicited play-like behaviour (Ellis and Wells, 2010). Griffith et al. (2000) investigated the potential for a synthetic feline facial pheromone to influence behaviour and food intake of sick cats in veterinary hospital environments – a potentially stressful environment for pet cats. They found that exposure to the synthetic feline facial pheromone was related to increased grooming, interest in food, and food intake when compared to cats exposed to a vehicle placebo. The authors interpreted these results as evidence of an anxiolytic effect of the synthetic feline facial pheromone.

Many studies of the effects of toys on domestic cats have been conducted. Toys seem like the most obvious way to enrich the environment of a seemingly bored individual. Animals not accustomed to toys may show an initial fear and avoidance of the item (Shimoji et al., 1993), followed by tentative exploration, and eventually play (Young, 2003). Animals more familiar with toys often respond by immediately exploring or playing with the object, or by showing other signs of excitement (Wood-Gush and Vestergaard, 1993). Although the exact function play behaviour serves is unclear (Young, 2003), it has been argued that is a desirable behaviour (Charmove and Anderson, 1989) and thus it has been proposed as a welfare indicator (Young, 2003). Denenberg (2003) attempted to determine toy preference in cats, by comparing the reaction of cats to 10 different toys of varying characteristics. Results indicated that cats have a strong preference for toys stimulating chase/predation and containing food (feeding enrichment discussed above), while the static or self-play toys were less preferred by cats. In his study however, the cats were stimulated with the toys by the author, and toys had varying degrees of human interaction. Thus, it is difficult to ascertain how much of the interest was a result of the enrichment of the toy, and how much was a result of the social enrichment. In a paper reporting the toy preferences of laboratory animals (including cats), Deluca and Kranda (1992) commented that no matter the species, a clear preference was exhibited for humans over the toys, although these findings were anecdotal only, and no scientific evaluation was carried out. Working with specific toys, de Monte and le Pape (1997) found that the introduction of toys – in the form of a ball and a log suspended to the side of the cage – in the enclosures of singly housed adult cats resulted in an increase of sniffing objects and play behaviour with objects, and a decrease of inactivity and self-play activities compared to a control period without either object. They also found that the duration of ball playing was greater than the duration of log playing, when log and ball were present simultaneously and separately, suggesting greater interest in the ball. Since inactivity was interpreted as detrimental to psychological well-being, the authors interpreted the introduction of toys – particularly the ball – as reducing boredom and thereby improving psychological well-being. The authors also recognised a loss of interest in the objects over time, which they noted is not observed in all species, suggesting a habituation effect and the importance of novelty. The habituation effect of cats to toys has been investigated more fully by Hall et al. (2002), who found that three sequential interactions with a toy was enough to cause almost complete habituation. They found that the amount of time between sessions influenced this effect, and that changing sensory characteristics of the toy elicited significant disinhibition, and renewed interest in the toy. They concluded that while object play is largely elicited by prey-like characteristics, it relies on change or novelty to be maintained.

In domestic cats, the effects of routine on urinary cortisol, hormone stimulation tests (ACTH and luteinising hormone releasing hormone (LHRH) – a tropic hormone released from the anterior pituitary which has a role in regulating reproduction), and behaviour were investigated by Carlstead et al. (1993). They found that cats subjected to irregular caretaking and feeding regimes exhibited significantly higher concentrations of urinary cortisol, enhanced sensitivity to ACTH, reduced sensitivity to LHRH, reduced frequency of active exploratory and play behaviour, and increased frequency of time awake/alert and attempting to hide. Furthermore, they found a negative correlation between frequency of hiding attempts and cortisol concentration, from which they suggested that hiding may represent an important strategy for coping with the stress of irregular routine.

It is also possible that individuality in the cat may affect their interaction with confinement, stress, and enrichment. McCune (1992) reported two different styles of behavioural responses to stress – passive and active – each of which are reported to respond differently to confinement (summarised in Ellis, 2009). It has been suggested that due to these different responses, different types of environmental enrichment may be more appropriate for these different types of responders (Ellis, 2009). However, further research is required to substantiate these claims.

It is important to note that behavioural problems – the most common being inappropriate waste elimination, including spraying and lack of litter box use – have been cited among the top reasons for relinquishing cats to shelters (summarised in Scarlett, 2008), yet it has been suggested that many of these problem behaviours could be addressed using environmental enrichment in the home (Overall and Dyer, 2005). While a further understanding of how environmental enrichment could be used in confined conditions to reduce anxiety and stress in shelters, perhaps an improved understanding and employment of these strategies in the home may help keep animals from ending up in shelters in the first place.

1.4.5 Asking the animals

It is clear that many types of environmental enrichment are provided to captive animals, but knowing what types will be the most biologically relevant or beneficial to a particular species or in a specific circumstance can be difficult. One way to answer this question is to simply ask the animal how they would allocate their time in a choice test.

Choice tests involve requiring individuals to choose between two or more different options or environments (Fraser and Matthews, 1997). These tests are best suited for simple tests to determine preference between similar options, such as level of light, type of substrate, or ambient temperature (summarised in Fraser and Matthews, 1997). However, questions are rarely as straight forwards as they initially seem, and researchers must take care to consider relevant factors, potentially confounding variables, contextual differences, behavioural experience, biological relevance, and individual differences. It is assumed that an animal's preference will correspond to choices that are best for the animal's welfare; however in some circumstances an animal may make a choice that is better for it in the short term, but worse for it in the long term. Results from choice tests must be interpreted with caution, as studying how an animal allocates its time may miss some very important factors. For example, if water and sleeping quarters are kept in different areas, an animal may be shown to use the sleeping quarters far more than the water, but obviously the water is of great importance to the animal. The behaviours associated with each resource take very different amounts of time, and thus the amount of time spent with each resource are very different. It is for this and other reasons that the simple choice test is widely criticized and its application can be controversial (summarised in Fraser and Matthews, 1997). However, the choice test can still give important results if designed carefully, especially when more complicated tests are impractical or too costly.

In an effort to resolve some of the issues surrounding assessing behavioural needs with a simple choice test, Dawkins (1983) introduced the idea of applying consumer-demand theory. Using this economic theory, researchers are able to quantify the behavioural needs of an animal by determining how much an animal is willing to 'pay' for access to a resource. In 2001, Mason et al. used this method to assess what resources were important to fur-farmed mink. The authors gave mink access to seven compartments – each containing different resources – and measured their consumer-demand for each resource. What differentiated this test from a simple choice test was that access to each compartment was controlled via a weighted flap door, and on each day the weight on the door increased. This meant that each day the mink would have to 'pay' a little more energy to access the resource. The thought behind this design was that as 'cost' for each resource increased, the amount the mink were willing to 'pay' to access it would become evident – meaning they would continue to 'pay' for resources that were truly important to their behavioural needs and would discontinue accessing resources that were considered less important. The authors found that mink would continue accessing a swimming pool compartment even after the weight applied to the door was greater than their body weight, while they no longer worked to gain access to some other compartments. In a second experiment in this paper, depriving mink of access to the swimming area resulted in an increase in urinary cortisol that was similar to the urinary cortisol increase in response to depriving mink of access to food, but far greater than the urinary cortisol response related to depriving access to other resources. These experiments make clear the benefit of applying consumer-demand theory to the study of animal welfare by quantifying motivation from an economic perspective, and solve some of the problems of simple choice tests.

1.5 Thesis objective

As the population of pet cats is growing (Rochlitz, 1999), so too is the unowned population. Although shelters have been suggested as potentially stressful environments for cats (Carlstead et al., 1993; Rochlitz, 1999), it is likely that millions pass through them every year – more than 100 000 cats pass through Canadian shelters alone (Canadian Federation of Humane Societies, 2010). Environmental enrichment provides a potential mechanism to reduce the stress of cats in these environments (Ellis, 2009), although investigation of this relationship using a range of indicators of stress is still needed. There is also a growing body of evidence that there is large variation of temperament and behavioural styles in cats (Mendl and Harcourt, 2000). It is likely that cats expressing different modes of a behavioural style would interact differently with enrichment items, and may derive benefit from them differentially. Considering this, the main objective of this thesis is to evaluate if environmental enrichment can alleviate stress due to caging in domestic cats, and examine the role of behavioural style in this relationship. This objective will be achieved using several sequential studies, each with their own methods, aims and hypotheses. These will be the content of Ch 3-8. Ch 9 will be a general discussion that will summarise the findings of all of the studies, and will present overall conclusions.

1.6 References

Altman, J., Wallace, R.B., Anderson, W.J., Das, G.D., 1968. Behaviourally induced changes in length of cerebrum in rats. Dev. Psychobiol. 1, 112-117.

Archer, J., 1979. Animals Under Stress. Edward Arnold, London.

Averill, J.R., 1973. Personal control over aversive stimuli and its relationship to stress. Psychol. Bull. 80, 286.

Baker, K.C., 2004. Benefits of positive human interaction for socially housed chimpanzees. Anim. Welfare 13, 239-245.

Belz, E.E., Kennell, J.S., Czambel, R.K., Rubin, R.T., Rhodes, M.E., 2003. Environmental enrichment lowers stress-responsive hormones in singly housed male and female rats. Pharmacology Biochemistry and Behavior 76, 481-486.

Bernstein, D.A., Nash, P.W., 2008. Essentials of Psychology, 4th ed. Houghton Mifflin Company, Boston.

Brambell, F.W.R., 1965. Report of the Technical Committee to Enquire into the Welfare of Animals Kept Under Intensive Livestock Husbandry Systems: Presented to Parliament by the Secretary of State for Scotland and the Minister of Agriculture, Fisheries and Food by Command of Her Majesty December, 1965. HM Stationery Office, London.

Broom, D.M., 1988. The scientific assessment of animal welfare. Appl. Anim. Behav. Sci. 20, 5-19.

Broom, D.M., Fraser, A.F., 2007. Domestic Animal Behaviour and Welfare, 4th ed. CABI, Wallingford.

Broom, D.M., Johnson, K.G., 1993. Stress and Animal Welfare. Chapman & Hall, London.

Buss, A.H., Plomin, R., 1975. A Temperament Theory of Personality Development. John Wiley & Sons Inc, Toronto.

Buss, A.H., Plomin, R., 1984. Temperament: Early Developing Personality Traits. Lawrence Erlbaum, Hillside, NJ.

Buss, A.H., Plomin, R., 1986. The EAS approach to temperament, in: Plomin, R., Dunn, J. (Eds.), The study of Temperament: Changes, Continuities, and Challenges. Lawrence Erlbaum, Hillsdale, NJ, pp. 67-79.

Cafazzo, S., Natoli, E., 2009. The social function of tail up in the domestic cat (*Felis silvestris catus*). Behav. Processes 80, 60-66.

Campbell, N.A., Reece, J.B., 2005. Biology, 7th ed. Pearson/Benjamin Cummings, New York, NY.

Canadian Animal Health Institute, 2012. Companion animal health [cited 2013, Oct 8]. Available from: http://www.cahi-icsa.ca/companion-animal-health/.

Canadian Federation of Humane Societies, 2010. Animal Shelter Statistics. [cited 2013, Nov 6]. Available from: http://cfhs.ca/athome/shelter_animal_statistics.

Carlstead, K., Brown, J.L., Strawn, W., 1993. Behavioral and Physiological Correlates of Stress in Laboratory Cats. Appl. Anim. Behav. Sci. 38, 143-158.

Carlstead, K., Shepherdson, D., 2000. Alleviating stress in zoo animals with environmental enrichment, in: Moberg, G.P., Mench, J.A. (Eds.) The Biology of Animal Stress. Basic Principles and Implications for Animal Welfare. CABI Publishing, New York, pp. 337-354.

Charmove, A.S., Anderson, J.R., 1989. Examining Environmental Enrichment, in: Segal, E.F. (Ed.) Housing, Care and Psychological Well Being of Captive and Laboratory Primates. Noyes Publications, Park Ridge, N.J., pp. 183-202.

Clarke, D.L., Wrigglesworth, D., Holmes, K., Hackett, R., Michel, K., 2005. Using environmental and feeding enrichment to facilitate feline weight loss. Journal of Animal Physiology & Animal Nutrition 89, 427-427.

Cloninger, C.R., 1994. Temperament and personality. Curr. Opin. Neurobiol. 4, 266-273.

Collard, R.R., 1967. Fear of strangers and play behaviour in kittens with varied social experience. Child Development 38, 877-891.

Cooper, C.L., Dewe, P., 2004. Stress :A Brief History. Blackwell Pub., Malden, MA.

Dantzer, R., 1989. Neuroendocrine correlates of control and coping., in: Steptoe, A., Appels, A. (Eds.) Stress, Personal Control and Health. John Wiley & Sons, Oxford, England, pp. 277-294.

Dawkins, M.S., 1983. Battery hens name their price: Consumer demand theory and the measurement of ethological 'needs'. Anim. Behav. 31, 1195-1205.

de Azevedo, C.S., Cipreste, C.F., Young, R.J., 2007. Environmental enrichment: a GAP analysis. Appl. Anim. Behav. Sci. 102, 329-343.

Deluca, A.M., Kranda, K.C., 1992. Environmental enrichment in a large animal facility. Lab Anim. 21, 38-44.

de Monte, M., LePape, G., 1997. Behavioural effects of cage enrichment in single-caged adult cats. Anim. Welfare 6, 53-66.

Denenberg, S., 2003. Cat toy play trial: A comparison of different toys. Proceedings of the Annual Scientific Symposium of Animal Behaviour, American Veterinary Society of Animal Behaviour, Denver Colorado.

Diamond, M.C., Law, F., Rhodes, H., Lindner, B., Rosenzweig, M.R., Krech, D., Bennett, E.L., 1966. Increases in cortical depth and glia numbers in rats subjected to enriched environment. J. Comp. Neurol. 128, 117-126.

Duncan, I.J.H., 1993. Welfare is to do with what animals feel. J. Agric. Environ. Ethics 6, 8-14.

Duncan, I.J.H., Fraser, D., 1997. Understanding Animal Welfare, in: Appleby, M.C., Hughes, B.O. (Eds.) Animal Welfare. CAB International, Cambridge, pp. 19-31.

Duncan, I., Olsson, I., 2001. Environmental enrichment: from flawed concept to pseudo-science. International Society of Applied Ethology 35th International Congress, 73.

Dunn, A.J., 1989. Psychoneuroimmunology for the psychoneuroendocrinologist: A review of animal studies of nervous system-immune system interactions. Psychoneuroendocrinology 14, 251-274.

Durr, R., Smith, C., 1997. Individual differences and their relation to social structure in domestic cats. Journal of Comparative Psychology 111, 412-418.

Edwards, D.S., Coyne, K., Dawson, S., Gaskell, R.M., Henley, W.E., Rogers, K., Wood, J.L.N., 2008. Risk factors for time to diagnosis of feline upper respiratory tract disease in UK animal adoption shelters. Prev. Vet. Med. 87, 327-339.

Ellis, S.L.H., 2009. Environmental enrichment: practical strategies for improving feline welfare. J. Feline Med. Surg. 11, 901-912.

Ellis, S.L.H., Wells, D.L., 2008. The influence of visual stimulation on the behaviour of cats housed in a rescue shelter. Appl. Anim. Behav. Sci. 113, 166-174.

Ellis, S.L.H., Wells, D.L., 2010. The influence of olfactory stimulation on the behaviour of cats housed in a rescue shelter. Appl. Anim. Behav. Sci. 123, 56-62.

Engh, A.L., Beehner, J.C., Bergman, T.J., Whitten, P.L., Hoffmeier, R.R., Seyfarth, R.M., Cheney, D.L., 2006. Female hierarchy instability, male immigration and infanticide increase glucocorticoid levels in female chacma baboons. Anim. Behav. 71, 1227-1237.

Feaver, J., Mendl, M., Bateson, P., 1986. A method for rating the individual distinctiveness of domestic cats. Anim. Behav. 34, 1016-1025.

Fogle, B., 1991. The Cat's Mind. Pelham, London.

Fraser, D., Matthews, L.R., 1997. Preference and motivation testing, in: Appleby, M.C., Hughes, B.O. (Eds.) Animal Welfare. CAB International, New York, pp. 159-173.

Fraser, D., 2008. Understanding Animal Welfare : The Science in its Cultural Context. Wiley-Blackwell, Ames, Iowa.

Gourkow, N., Fraser, D., 2006. The effect of housing and handling practices on the welfare, behaviour and selection of domestic cats (*Felis sylvestris catus*) by adopters in an animal shelter. Anim. Welfare 15, 371-377.

Griffith, C.A., Steigerwald, E.S., Buffington, C.A.T., 2000. Effects of a synthetic facial pheromone on behavior of cats. J. Am. Vet. Med. Assoc. 217, 1154-1156.

Hall, S.L., Bradshaw, J.W.S., Robinson, I.H., 2002. Object play in adult domestic cats: the roles of habituation and disinhibition. Appl. Anim. Behav. Sci. 79, 263-271.

Hart, B.L., Hart, L.A., 1984. Selecting the best companion animal: breed and gender behavioural profiles, in: Anderson, R.K., Hart, B.L., Hart, L.A. (Eds.) The Pet Connection. Its Influence on Our Health. University of Minnesota Press, Minneapolis, pp. 180-193.

Hawkins, K.R., Bradshaw, J.W.S., Casey, R.A., 2004. Correlating cortisol with a behavioural measure of stress in rescue shelter cats. Anim. Welfare 13, S242-S243.

Hennessy, M.B., Voith, V.L., Mazzei, S.J., Buttram, J., Miller, D.D., Linden, F., 2001. Behavior and cortisol levels of dogs in a public animal shelter, and an exploration of the ability of these measures to predict problem behavior after adoption. Appl. Anim. Behav. Sci. 73, 217-233.

Henry, J.P., Stephens, P.M., 1977. Stress, Health, and the Social Environment: A Sociobiologic Approach to Medicine. Springer-Verlag, New York.

Hodges, J.K., Heistermann, M., 2003. Field Endocrinology: Monitoring Hormonal Changes in Free-Ranging Primates, in: Setchell, J.M., Curtis, D.J. (Eds.) Field and Laboratory Methods in Primatology. A Practical Guide. Cambridge University Press, Cambridge, pp. 353-370.

Hughes, B.O., Duncan, I.J.H., 1988. The notion of ethological 'need', models of motivation and animal welfare. Anim. Behav. 36, 1696-1707.

Iki, T., Ahrens, F., Pasche, K.H., Bartels, A., Erhard, M.H., 2011. Relationships between scores of the feline temperament profile and behavioural and adrenocortical responses to a mild stressor in cats. Appl. Anim. Behav. Sci. 132, 71-80.

Jones, A.C., Gosling, S.D., 2005. Temperament and personality in dogs (*Canis familiaris*): A review and evaluation of past research. Appl. Anim. Behav. Sci. 95, 1-53.

Jones, E., Horton, B.J., 1984. Gene frequencies and body weight of cats, *Felis catus* (L.), from five australian localities and from Macquarie Island. Aust. J. Zool. 32, 231-237.

Karsh, E.B., 1984. Factors influencing the socialization of cats to people. , 207-215.

Karsh, E.B., Turner, D.C., 1988. The human-cat relationship, in: Turner, D.C., Bateson, P. (Eds.) The Domestic Cat: The Biology of its Behaviour. Cambridge University Press, New York, NY, pp. 159-177.

Keeling, L., Jensen, P., 2009. Abnormal behaviour, stress and welfare, in: Jensen, P. (Ed.), 2 ed. Cabi, Cambridge, MA, pp. 85-101.

Kessler, M.R., Turner, D.C., 1997. Stress and adaptation of cats (*Felis silvestris catus*) housed singly, in pairs and in groups in boarding catteries. Anim. Welfare 6, 243-254.

Kessler, M.R., Turner, D.C., 1999. Socialization and stress in cats (*Felis silvestris catus*) housed singly and in groups in animal shelters. Anim. Welfare 8, 15-26.

Kry, K., Casey, R., 2007. The effect of hiding enrichment on stress levels and behaviour of domestic cats (*Felis sylvestris catus*) in a shelter setting and the implications for adoption potential. Anim. Welfare 16, 375-383.

Larsson, F., Winblad, B., Mohammed, A.H., 2002. Psychological stress and environmental adaptation in enriched vs. impoverished housed rats. Pharmacology Biochemistry and Behavior 73, 193-207.

Lee, R.L., Zeglen, M.E., Ryan, T., Hines, L.M., 1983. Guidelines: Animals in nursing homes. California Veterinarian 3, 22a-26a.

Lowe, S.E., Bradshaw, J.W.S., 2001. Ontogeny of individuality in the domestic cat in the home environment. Anim. Behav. 61, 231-237.

Lowe, S.E., Bradshaw, J.W.S., 2002. Responses of pet cats to being held by an unfamiliar person, from weaning to three years of age. Anthrozoos 15, 69-79.

Lyons, D.M., 1989. Individual differences in temperament of dairy goats and the inhibition of milk ejection. Appl. Anim. Behav. Sci. 22, 269-282.

Manteca, X., Deag, J.M., 1993. Individual differences in temperament of domestic animals: a review of methodology. Anim. Welfare 2, 247-268.

Markowitz, H., Aday, C., 1998. Power for captive animals: contingencies and nature, in: Shepherdson, D.J., Mellen, J.D., Hutchins, M. (Eds.) Second Nature: Environmental Enrichment for Captive Animals. Smithsonian Institution Press, Washington, pp. 47-58.

Martrenchar, A., Huonnic, D., Cotte, J.P., 2001. Influence of environmental enrichment on injurious pecking and perching behaviour in young turkeys. Br. Poult. Sci. 42, 161-170.

Mason, G.J., 2010. Species differences in responses to captivity: stress, welfare and the comparative method. Trends in ecology & evolution 25, 713-721.

Mason, J.W., 1971. A re-evaluation of the concept of 'non-specificity'in stress theory. J. Psychiatr. Res. 8, 323-333.

Mason, G., Cooper, J., Clarebrough, C., 2001. Frustrations of fur-farmed mink. Nature 410, 35-36.

Matteri, R.L., Carrol, J.A., Dyer, C.J., 2000. Neuroendrocrine Responses to Stress, in: Moberg, G.P., Mench, J.A. (Eds.) The Biology of Animal Stress: Basic Principles and Implications for Animal Welfare. CAB International, New York, NY, pp. 43-76.

McCune, S., 1995. The impact of paternity and early socialization on the development of cats behavior to people and novel objects. Appl. Anim. Behav. Sci. 45, 109-124.

McCune, S., 1992. Temperament and the welfare of caged cats. Ph.D. Thesis, University of Cambridge.

Meier, M., Turner, D.C., 1985. Reactions of house cats during encounters with a strange person: Evidence for two personality types. Journal of the Delta Society 2, 45-53.

Mellen, J.D., 1992. Effects of early rearing experience in subsequent adult sexual behaviour using domestic cats (*Felis catus*) as a model for exotic small felids. Zoo Biology 11, 17-32.

Mendl, M.T., 1986. Effects of Litter Size and Sex of Young on Behavioural Development in Domestic Cats. Ph.D. Thesis, University of Cambridge.

Mendl, M., Harcourt, R., 2000. Individuality in the cat: origins, development and stability, in: Turner, D.C., Bateson, P. (Eds.) The Domestic Cat: The Biology of its Behaviour, 2nd ed. Cambridge University Press, Cambridge, pp. 47-64.

Mertens, C., Turner, D.C., 1989. Experimental analysis of human-cat interactions during first encounters. Anthrozoös 2, 83-97.

Moberg, G.P., 2000. Biological Response to Stress: Implications for Animal Welfare, in: Moberg, G.P., Mench, J.A. (Eds.) The Biology of Animal Stress: Basic Principles and Implications for Animal Welfare. CABI Pub., New York, NY, pp. 1-22.

Moberg, G.P., 1985. Biological response to stress: key to assessment of animal well-being?, in: Moberg, G.P. (Ed.) Animal Stress. American Physiological Society, Bethesda, Md., pp. 29-49.

Monat, A., Lazarus, R.S., 1991. Stress and Coping: An Anthology, 3 ed. Columbia University Press, New York.

Moncek, F., Duncko, R., Johansson, B.B., Jezova, D., 2004. Effect of Environmental Enrichment on Stress Related Systems in Rats. J. Neuroendocrinol. 16, 423-431.

Morgan, K.N., Tromborg, C.T., 2007. Sources of stress in captivity. Appl. Anim. Behav. Sci. 102, 262-302.

Mormède, P., Andanson, S., Aupérin, B., Beerda, B., Guémené, D., Malmkvist, J., Manteca, X., Manteuffel, G., Prunet, P., van Reenen, C.G., Richard, S., Veissier, I., 2007. Exploration of the hypothalamic-pituitary-adrenal function as a tool to evaluate animal welfare. Physiol. Behav. 92, 317-339.

Naka, F., Shiga, T., Yaguchi, M., Okado, N., 2002. An enriched environment increases noradrenaline concentration in the mouse brain. Brain Res. 924, 124-126.

Newberry, R.C., 1995. Environmental enrichment: Increasing the biological relevance of captive environments. Appl. Anim. Behav. Sci. 44, 229-243.

Overall, K.L., Dyer, D., 2005. Enrichment strategies for laboratory animals from the viewpoint of clinical veterinary behavioral medicine: Emphasis on cats and dogs. Ilar Journal 46, 202-216.

Oxford University Press, 2000. A Dictionary of Biology, 4th ed. Martin, E., Hines R., editors. Oxford University Press: Oxford England.

Pacteau, C., Einon, D., Sinden, J., 1989. Early rearing environment and dorsal hippocampal ibotenic acid lesions: long-term influences on spatial learning and altenation in the rat. Behav. Brain Res. 34, 79-96.

Palme, R., 2012. Monitoring stress hormone metabolites as a useful, non-invasive tool for welfare assessment in farm animals. Anim. Welfare 21, 331.

Pontier, D., Rioux, N., Heizmann, A., 1995. Evidence of selection on the orange allele in the domestic cat *Felis catus*: The role of social structure. Oikos 73, pp. 299-308.

Réale, D., Reader, S.M., Sol, D., McDougall, P.T., Dingemanse, N.J., 2007. Integrating animal temperament within ecology and evolution. Biol. Rev. 82, 291-318.

Reed, H.J., Wilkins, L.J., Austin, S.D., Gregory, N.G., 1993. The effect of environmental enrichment during rearing on fear reactions and depopulation trauma in adult caged hens. Appl. Anim. Behav. Sci. 36, 39-46.

Reisner, I.R., Houpt, K.A., Erb, H.N., Quimby, F.W., 1994. Friendliness to humans and defensive aggression in cats: the influence of handling and paternity. Physiol. Behav. 55, 1119-1124.

Rochlitz, I., 1999. Recommendations for the housing of cats in the home, in catteries and animal shelters, in laboratories and in veterinary surgeries. J. Feline Med. Surg. 1, 181-91.

Rochlitz, I., Podberscek, A., Broom, D., 1998. Welfare of cats in a quarantine cattery. Vet. Rec. 143, 35-39.

Rochlitz, I., 2000. Feline welfare issues, in: Turner, D.C., Bateson, P. (Eds.) The Domestic Cat: The Biology of its Behaviour, 2nd ed. Cambridge University Press, Cambridge, pp. 207-226.

Rosenzweig, M.R., Bennett, E.L., 1969. Effects of differential environments on brain weights and enzyme activities in gerbils, rats, and mice. Dev. Psychobiol. 2, 87-95.

Rothbart, M.K., Derryberry, D., 1981. Development of individual differences in temperament, in: Lamb, M.E., Brown, A.L. (Eds.) Advances in Developmental Psychology, Vol. 1. Erlbaum, Hilsdale, NJ, pp. 37-86.

Russell, W.M.S. and Burch, R.L., (1959). The Principles of Humane Experimental Technique. Methuen, London.

Scarlett, J.M., 2008. Interface of epidemiology, pet population issues and policy. Prev. Vet. Med. 86, 188-197.

Schapiro, S.J., Nehete, P.N., Perlman, J.E., Sastry, K.J., 2000. A comparison of cell-mediated immune responses in rhesus macaques housed singly, in pairs, or in groups. Appl. Anim. Behav. Sci. 68, 67-84.

Schipper, L.L., Vinke, C.M., Schilder, M.B.H., Spruijt, B.M., 2008. The effect of feeding enrichment toys on the behaviour of kennelled dogs (*Canis familiaris*). Appl. Anim. Behav. Sci. 114, 182-195.

Selye, H., 1936. A Syndrome Produced by Diverse Nocuous Agents. Nature 138, 32.

Selye, H., 1974. Stress without Distress, 1 ed. Lippincott, Philadelphia.

Selye, H., 1976. The Stress of Life, Revis ed. McGraw-Hill, New York.

Serpell, J.A., Hsu, Y., 2001. Development and validation of a novel method for evaluating behavior and temperament in guide dogs. Appl. Anim. Behav. Sci. 72, 347-364.

Shepherdson, D.J., 1998. Tracing the path of environmental enrichment in zoos, in: Shepherdson, D.J., Mellen, J.D., Hutchins, M. (Eds.) Second Nature: Environmental Enrichment for Captive Animals,. Smithsonian Institution Press, Washington, pp. 1-12.

Shimoji, M., Bowers, C.L., Crockett, C.M., 1993. Initial response to the introduction of a PVC perch by singly caged *Macaca fascicularis*. Laboratory Primate Newsletter 32, 8-11.

Shyne, A., 2006. Meta-Analytic Review of the Effects of Enrichment on Stereotypic Behavior in Zoo Mammals. Zoo Biol. 25, 317-337.

Siegford, J.M., Walshaw, S.O., Brunner, P., Zanella, A.J., 2003. Validation of a temperament test for domestic cats. Anthrozoos 16, 332-351.

Skerrett, J., 2011. 50,000 feral cats in the city. Winnipeg Free Press. 2011, April 3; A1.

Slabbert, J., Odendaal, J., 1999. Early prediction of adult police dog efficiency—a longitudinal study. Appl. Anim. Behav. Sci. 64, 269-288.

Smith, D.F., Durman, K.J., Roy, D.B., Bradshaw, J.W., 1994. Behavioural aspects of the welfare of rescued cats. The Journal of the Feline Advisory Bureau 31, 25-28.

Stella, J., Croney, C., Buffington, T., 2013. Effects of stressors on the behavior and physiology of domestic cats. Appl. Anim. Behav. Sci. 143, 157-163.

Stevenson-Hinde, J., 1983. Individual characteristics: a statement of a problem, in: Hinde, R.A. (Ed.) Primate Social Relationships. Blackwell Scientific Publications, Oxford, pp. 28-35.

Tanaka, A., Wagner, D.C., Kass, P.H., Hurley, K.F., 2012. Associations among weight loss, stress, and upper respiratory tract infection in shelter cats. J. Am. Vet. Med. Assoc. 240, 570-576.

The Toronto Humane Society, Feral Cat Awareness. 2011. [cited 2013, Oct 8]. Available from: http://www.torontohumanesociety.com/feral_cat_awareness.htm.

Todd, N.B., 1977. Cats and commerce. Scientific American 237, 100-107.

Touma, C., Palme, R., 2005. Measuring fecal glucocorticoid metabolites in mammals and birds: the importance of validation. Ann. N. Y. Acad. Sci. 1046, 54-74.

Turner, D.C., 2000a. Human-cat interactions: relationships with, and breed differences between, non-pedigree, Persian and Siamese Cats, in: Podberscek, A.L., Paul, E.S., Serpell, J.A. (Eds.). Cambridge University Press, Cambridge, pp. 257-271.

Turner, D.C., 2000b. The human-cat relationship, in: Turner, D.C., Bateson, P. (Eds.), 2nd ed. Cambridge University Press, Cambridge, pp. 193-206.

Turner, D.C., Feaver, J., Mendl, M., Bateson, P., 1986. Variation in domestic cat behaviour towards humans: A paternal effect. Anim. Behav. 34, 1890-1892.

van Praag, H., Kempermann, G., Gage, F.H., 2000. Neural Consequences of Environmental Enrichment. Nature Reviews Neuroscience 1, 191-198.

Veissier, I., Boissy, A., 2007. Stress and welfare: two complementary concepts that are intrinsically related to the animal's point of view. Physiol. Behav. 92, 429-433.

Weiss, E., Miller, K., Heather Mohan-Gibbons, Vela, C., 2012. Why Did You Choose This Pet?: Adopters and Pet Selection Preferences in Five Animal Shelters in the United States. Animals, 144.

Wells, D.L., Irwin, R.M., 2008. Auditory stimulation as enrichment for zoo-housed Asian elephants (*Elephas maximus*). Anim. Welfare 17, 335-340.

Wells, D.L., 2004. A review of environmental enrichment for kennelled dogs, *Canis familiaris*. Appl. Anim. Behav. Sci. 85, 307-317.

Wemelsfelder, F., 1997. Life in captivity: its lack of opportunities for variable behaviour. Appl. Anim. Behav. Sci. 54, 67-70.

Wenstrup, J., Dowidchuk, A., 1999. Pet Overpopulation: Data and Measurement Issues in Shelters. Journal of Applied Animal Welfare Science 2, 303.

Whitten, P.L., Brockman, D.K., 1998. Recent Advances in Noninvasive Techniques to Monitor Hormone-Behavior Interactions. Yearbook of Physical Anthropology 41, 1.

Wilson, M., Warren, J.M., Abbott, L., 1965. Infantile stimulation, activity, and learning by cats. Child Devel. 36, 843-853.

Wood-Gush, D.G.M., Vestergaard, K., 1989. Exploratory behavior and the welfare of intensively kept animals. J. Agric. Ethics, 2, 161-169.

Wood-Gush, D.G.M., Vestergaard, K., 1993. Inquisitive exploration in pigs. Anim. Behav. 45, 185-187.

Young, R.J., 2003. Environmental Enrichment for Captive Animals. Blackwell Science, Malden, MA.

2 AGREEMENT AND RELIABILITY FOR OBSERVATIONAL DATA

2.1 Introduction

Throughout this thesis there are many instances in which two or more sets of data need to be assessed for agreement or reliability. Often in the literature, reporting of agreement and reliability is incomplete or inappropriate (Kottner et al., 2011). Inconsistent methods are used across studies (Gisev et al., 2013), methods are used that have been shown to be inadequate (i.e. use of Pearson's correlation coefficient to assess agreement, as cited in Bland and Altman, 1986) or authors are unclear about exactly what methods they used and/or how they have been applied (Kottner et al., 2011). These problems are often pronounced in behavioural studies, because specifics for applying these statistical techniques to observational behavioural data can be ambiguous. Explanations of which methods to use in which situations – and how to apply or interpret them – are inadequate in the animal welfare/behaviour specific literature. The purpose of this chapter is to outline the appropriate assessments, and explain how they have been applied in this thesis.

One issue central to the reporting problems, is that often the terms agreement and reliability are used interchangeably, when in fact they describe different concepts. Agreement assesses the similarity of two or more sets of measurements (de Vet, 2005). Reliability coefficients assess how well subjects or objects can be distinguished from others in the data set (de Vet et al., 2006). Although the distinction between these terms may not seem intuitive at first, it is possible for two sets of measurements to have high agreement but low reliability, and vice versa. For example, if two observers agree on the percentage of time each subject spends sleeping, the data would have perfect agreement. However, if all of the subjects spent the same percentage of time sleeping, the data would have low reliability. This is a slightly more narrow definition of reliability than is traditionally used in behavioural science; for example the terms inter-observer reliability and inter-observer agreement are commonly used to refer to the same concept. However, the specific statistical definitions suggested above rely on ratios of variance, specifically, the variance between subjects divided by the total variance (discussed further below).

There are different measures used to assess agreement and reliability, but choice of measure also depends on the nature of the variable (binary/categorical, ordinal, or continuous data) and the number of observers (for measures of agreement only). Table 2.1 outlines the appropriate measures to use to assess both agreement and reliability in continuous or categorical data, with two or more than two raters.

Table 2.1 Recommended measures of agreement and reliability for different types of data and number of raters

		Agreement	Reliability
Binary/ categorical	2 observers	Percentage of agreement	NA
		Cohen's kappa	
	> 2 observers	Percentage of agreement	
		Fleiss's kappa	
Ordinal	2 observers	Weighted kappa	ICC†
	> 2 observers	Kendall's coefficient of correlation	
Continuous	2 observers	Limits of agreement	ICC
		Concordance correlation coefficient	
	> 2 observers	Overall concordance correlation coefficient‡	

NA=Not applicable – reliability cannot usually be computed for categorical data
ICC=Intraclass correlation coefficient
†if the scale assumes continuous properties
‡Described by Barnhart et al. (2002), but its use has yet to become standard.

2.2 Measures of agreement

2.2.1 Categorical and ordinal data

Percentage of agreement

Perhaps the most common measure reported is the percentage of agreement. This measure divides all instances in which the observers agree with the total number of possible instances, and the formula is given in Figure 2.1. This measure can be calculated for either two or more than two 2 raters, but it is negatively related to the number of raters, as only perfect agreement between all raters would be reflected. Percentage of agreement is easy to calculate and interpret, but there are two associated shortcomings. First, it is insensitive to degrees of agreement, meaning that even if all scores or ratings were very close to identical (but not exactly), they would have the same percentage of agreement as if all scores or ratings were drastically different (Mitchell, 1979). Second, percentage of agreement includes agreement due to chance. Some behaviours – especially those exhibited at extremes – are especially vulnerable to conflated agreement due to chance (Mitchell, 1979). In these circumstances, percent of agreement overestimates the real agreement.

$$\textbf{Percentage of agreement} = \frac{\textbf{Number of identical responses}}{\textbf{Total number of responses}} \times \mathbf{100}$$

$$\textbf{ICC} = \frac{\textbf{Between subject variance}}{\textbf{Between subject variance } + \textbf{Within subject variance}}$$

$$\textbf{Kappa} = \frac{\textbf{Prop. observed agreement} - \textbf{Prop. agreement expected due to chance}}{\mathbf{1} - \textbf{Prop. agreement expected due to chance}}$$

$$\textbf{Cell weighting (linear)} = \mathbf{1} - \frac{\textbf{Distance from agreement}}{\textbf{Maximum possible distance from agreement}}$$

Figure 2.1 Formulas for percentage of agreement, kappa, cell weighting for linear weighted kappa, and ICC

Kappa

Kappa attempts to correct the findings of the percentage agreement measure to account for agreements due to chance alone. There are a number of different kappa indices for use in different circumstances; these are described below. For all kappa indices, values range between -1 and +1, where +1 indicates perfect agreement, 0 indicates agreement due to chance alone, and a negative value indicates agreement is less than expected by chance (although values <0 are rare). Interpretation of kappa values can be problematic, and although many authors recommend against using standard rules of acceptability these are widely employed; the most popular being those of Landis and Koch (1977), which are given in Figure 2.2. In addition, it is important that standard errors are calculated and interpretations of values take these into consideration (Gisev et al., 2013). There are three assumptions implicit in kappa: 1) the subjects being rated must be independent; 2) the raters must be independent; and 3) the classes must be mutually exclusive, and exhaustive (Tinsley and Weiss, 2000). Although kappa is superior to percentage of agreement in that it corrects for chance agreement, there are a few areas in which kappa has flaws. First, kappa is dependent on the number of classes (or behaviours) being assessed. A high number of classes can often result in a lower kappa. Second, kappa is dependent on prevalence of the classes. A high prevalence can contribute to high expected agreement, minimising the impact of actual agreement. Third, kappa is dependent on bias, or, the distribution of the agreement between classes. If the agreement between the observers is unbalanced (i.e. if observers always agree when the subject is resting or walking, but rarely agree if the subject is jogging or running) this can result in a higher kappa than if the same percentage of agreement was evenly distributed among the classes.

Cohen's kappa

Cohen's kappa is primarily used with nominal categorical variables assessed by two raters. It is defined as the proportion of possible agreement beyond chance that was actually achieved (Dohoo et al., 2009), and the formula is given in Figure 2.1. Like the percentage of agreement measure, Cohen's kappa is insensitive to degrees of agreement.

<table>
<tr><th></th><th>0.0</th><th>0.1</th><th>0.2</th><th>0.3</th><th>0.4</th><th>0.5</th><th>0.6</th><th>0.7</th><th>0.8</th><th>0.9 1.0</th></tr>
<tr><td>Proportion of agreement (Cooper et al., 2007)</td><td colspan="8">Not acceptable</td><td colspan="2">Acceptable</td></tr>
<tr><td>Kappa (Landis and Koch, 1977)</td><td colspan="2">No agreement</td><td colspan="2">Slight agreement</td><td colspan="2">Moderate agreement</td><td colspan="2">Substantial agreement</td><td colspan="2">Almost perfect agreement</td></tr>
<tr><td>Correlation coefficients (Cohen, 1988)</td><td colspan="3">Small effect size</td><td colspan="2">Medium effect size</td><td colspan="5">Large effect size</td></tr>
<tr><td>Correlation coefficients (Sprinthall and Fisk, 2002)</td><td colspan="2">Slight correlation</td><td colspan="2">Low correlation</td><td colspan="3">Moderate correlation</td><td colspan="2">High correlation</td><td>Very high correlation</td></tr>
<tr><td>Correlation coefficients (Ferguson, 2009)</td><td colspan="2"></td><td colspan="2">Recommended minimum effect size</td><td colspan="4">Moderate effect</td><td colspan="2">Strong effect</td></tr>
<tr><td>ICC (Gibbons et al., 2009)</td><td colspan="5">Low reliability
(>50% of variance within subject)</td><td colspan="5">High reliability
(>50% between subjects)</td></tr>
</table>

Figure 2.2 Interpretations of selected agreement and reliability measures suggested by various authors. These interpretations are guidelines only and the use of such standardised cut-offs is debatable, as different scenarios may require more or less stringent levels of agreement and/or reliability

Weighted kappa

This measure is used with ordinal variables assessed by two raters, and is sensitive to degrees of agreement. This means that there is room to acknowledge instances in which raters did not agree perfectly as a partial agreement, by assigning less weight as to categories further apart (de Vet, 2005). For example, if two raters have to assess friendliness in animals on a scale of 1-10, ratings of 8 and 9 by respective raters would be given a higher weighting than ratings of 1 and 9. Due to the consideration of partial agreements, weighted kappa tends to be higher than Cohen's kappa. The formula for linear cell weighting is given in Figure 2.1.

Fleiss's kappa

When more than two raters are assessing the classes, a Fleiss's kappa is utilised. Calculation of Fleiss's kappa is only possible with categorical data (either nominal or ordinal). This measure does not assume all classes have been assessed by the same raters (Gisev et al., 2013).

Kendall's coefficient of concordance

The Kendall's coefficient of concordance (W) is used when multiple raters are assessing ordinal variables. Unlike the kappa family of indices, values for W range from 0-1, because it is impossible to achieve complete disagreement with more than 2 raters. Although Landis and Koch's levels of interpretation were originally proposed for kappa values, they have been extended to interpret W values. While this measure can theoretically be used with any number of raters, the likelihood of W values being high reduces as the number of raters increases (Gisev et al., 2013).

2.2.2 Continuous data

Pearson's product-moment correlation coefficient

Some authors use Pearson's product-moment correlation coefficient (r) to assess the agreement of two sets of continuous data, as it measures the strength of the linear relationship between two variables (Sheskin, 2003). r can range from -1 to +1, with -1 representing a perfect negative relationship (as one variable decreases, the other increases) , +1 representing a perfect positive relationship (as one variable increases, so does the other), and 0 representing no linear relationship. However, two data sets can have a perfect linear relationship (either positive or negative) without agreement. For example, if two raters are asked to assess four subjects for their percentage of time spent grooming daily, it is possible that the first rater would report the percentages 10, 20, 30, and 40 while the second rater would assign the scores 30, 40, 50, and 60. In this example the r would be 1, but there would have been no instance in which the raters actually agreed. Rater 1 always tended to assign scores lower than rater 2. This is called observer bias, and is a potential pitfall that this measure is not equipped to identify. Furthermore, inference for r requires normally distributed data, as results can be strongly influenced by outliers and skewness.

Spearman's nonparametric correlation coefficient

Spearman's correlation coefficient (r_s) also measures the strength of the linear relationship between two continuous sets of data, after ranking the subjects from one to N based on their values, where N equals the number of subjects (Weinberg and Abramowitz, 2002). After replacing values with their relative ranks, r_s has the same range as r, but has no distributional assumptions or sensitivity to outliers. Although this method is also susceptible to observer bias, its usage in assessment of agreement may be slightly more acceptable than r in some circumstances. For example, while the magnitude of individual responses can change over time, the rank-order consistency of individuals in a group could stay relatively stable (Roberts and DelVecchio, 2000) in terms of the exhibition of temperament traits/behavioural styles.

Concordance correlation coefficient

The concordance correlation coefficient (CCC) is also used when two raters are assessing continuous data. This measure builds on r, but includes a bias correction factor that measures how far the line of best fit deviates from the 45^{o} line of origin (that is, y=x) allowing it to account for accuracy and precision (Lin, 1989). Its output and interpretation is similar to those of r, and it assumes a normal distribution (Carrasco and Jove, 2003). An overall concordance correlation coefficient has also been proposed for use with more than two observers (Barnhart et al., 2002), but its use has yet to become standard.

Limits of Agreements

This procedure (also called a Bland-Altman plot) presents an alternative technique for assessing continuous data with two raters. It uses a graphical approach to distinguish between random and non-random errors, and plots the difference between the rater's scores (x) against their mean value (y) (Bland and Altman, 1986). Visual inspection of the plot will then reveal the range of differences. Positive and negative limits of agreement are then established at -2 and +2 standard deviations, which should contain 95% of the differences for the two raters to be considered in agreement (de Vet, 2005).

2.3 Measures of reliability

2.3.1 Intraclass correlation coefficient

The intraclass correlation coefficient (ICC) can be used to assess the ability of two or more raters to discriminate between the subjects being rated, on a continuous or ordinal scale (assuming the ordinal scale assumes continuous properties). The ICC is based on an ANOVA framework, and thus assumes that the residuals have a normal distribution (Kottner et al., 2011). One may distinguish between up to 6 different ICCs, as described by Shrout and Fleiss (1979). Often authors fail to state which ICC has been used (Kottner et al., 2011), however the most common ICC formula (and the one used throughout this thesis) evaluates if there is more variance between subjects or within them, and is given in Figure 2.1. It has been shown (Carrasco and Jove, 2003; Chen and Barnhart, 2008) that the results of the CCC are almost identical to those of the ICC in many circumstances, as they are both estimated through variance. The ICC ranges from 0 to 1. If there is no variability within subjects and there is variability between subjects, the ICC is 1. If the measurement error is equal to the variability between subjects, the ICC is 0.5. If there is no variability between subjects and there is variability within subjects, the ICC is 0. Thus, ICC is affected by the heterogeneity of the population, with lower ICCs for homogenous populations (Mitchell, 1979). This is unlike most measures of agreement, which would be likely to be highest in homogeneous groups.

2.4 Significance tests

To assess if data collected by two observers watching the same subjects at the same time agree, it can be tempting to use a paired t-test (or for binary data, McNemar's test). Using a significance test such as this may seem attractive because the results require little interpretation; they are either significant or they are not. However, the hypothesis tested is if there is a systematic difference between the values, with no regard to whether there is random difference. In a practical sense, this investigates if the values from one data set are consistently higher/lower than those from the other data set. While non-significant results would mean that there is no reason to believe one observer systematically gives higher values, this makes no statement about the actual agreement of the data.

In this thesis there are places where a two-way ANOVA with cat as a random factor (for parametric analysis) or a Friedman's test (for non-parametric analysis) are utilised to assess concepts akin to agreement and reliability. In these instances, one observer is collecting data on behaviours exhibited by a group of animals at a number of time-points, and the author would like to ensure that there is no systematic difference in the exhibition of these behaviours between any of the time points. The lack of differences across time is similar to showing the behaviours exhibited at each time point agree, but with a more specific hypothesis. This makes analysis with a significance test sensible. However, the researcher would also like to show that there is difference in the exhibition of the behaviours between subjects, proving that these behaviours could be used to categorise subjects into different groups, similar to reliability. However, this would not give evidence that these subjects would fall under the same category at each time-point. Therefore, in order to support the findings of these analyses, an ICC was computed for the data that met the assumptions of the ANOVA model. Unfortunately, there was no measure of reliability available to assess comparable data for which the assumption of normality could not be met.

One of the drawbacks of using a significance test to show that exhibition of a behavior is not significantly different across time-points is the influence of sample size. The larger the sample size, the more likely a small difference across time-points (which may not be biologically significant or relevant practically) is to become statistically significant, and thus the measures are rejected as stable across time. Therefore, in larger data sets using an agreement or reliability measure with a set cut-point of 'acceptability' is more appropriate, as this does not depend on the amount of data. However in smaller data sets significance testing may be more useful because it is less likely to find a significant difference between variables with very small biological differences.

2.5 Interpreting measures of agreement and reliability

Agreement and/or reliability analyses are often conducted in one of two scenarios: 1) to assess the repeatability of a method of collecting/analysing data, and 2) to assess the consistency of exhibition of behaviour by a subject or group of subjects over time. The first scenario compares measurements collected from the same subject and time period either by different individuals (inter-observer), by one individual (perhaps scoring a video tape) twice (intra-observer), or by tools/tests designed to reflect behaviours exhibited (method comparison). Alternatively, the second scenario compares measurements collected from the same subject, in similar conditions, but at different times. In this scenario, data are most commonly collected by one observer assessing the behaviours exhibited by a group of subjects at different time-points, but in similar conditions. Although guidelines for interpretation of different agreement and reliability measures have been given in Figure 2.2, it is logical to assume that the agreement and reliability of data from scenario one are likely to be higher than those from data collected in scenario two, as the former is measuring the exact same occurrence, while the latter is comparing an animal's behaviour under similar circumstances. The choice of the cut-point representing an 'acceptable' level of agreement or reliability should be specific to the nature of the data, and it is important to keep the difference between scenarios one and two in mind when making this decision.

2.6 Agreement statistics calculated by Observer

In behavioural research, it is common to find observational data analysed using Noldus Observer software (Noldus Information Technology, Wageningen, The Netherlands). This program can be used to record either event or state behaviours, or a combination of the two. Data resulting from analysis of strictly event behaviours would consist of the number of times a behaviour was exhibited within the time-window. Data resulting from analysis of strictly state behaviours essentially produces a time-budget of how the subject allocated their time between the different behaviours defined in the ethogram (whether they be postures, locations, activities, facial expressions, etc), and can be expressed in a number of ways: 1) frequency each behaviour was exhibited within the time-window; 2) the duration of each behaviour; and 3) with a little manipulation, the percentage of the time-window that the subject was engaged in each behaviour. In this thesis, all quantitative behaviours have been recorded as state behaviours, whether expressed as frequencies, durations, or percentages.

2.6.1 Application of measures of agreement

Noldus Observer version 5.0 or later contain an analysis feature which calculates the agreement between two data sets, producing the parameters Cohen's kappa, r, and percentage agreement (it also produces the index of concordance, but this is simply the percentage of agreement expressed as a proportion). Jansen et al. (2003) state that Cohen's kappa is for use with nominal data, r is for use with continuous data, and percentage agreement is for use with data that has no measurement scale, but it can be difficult knowing how to fit time-budget data into these categories. Frequency, duration, and percentage are all continuous variables, however, expressing their distribution across behaviors exhibited during a specific period of time allows the use of calculations originally developed for categorical variables.

Observer can calculate the three measures of agreement in two different ways: on overall totals within the time-window, or by taking the temporal sequence of the behaviours into account.

While the user's manual (Noldus Information Technology, 2003) calls these continuous and nominal scales respectively, the terms tallied and sequenced data will be used here, in order to avoid confusion with the way the words continuous and nominal are used above. The differences between these two ways of analysing the agreement can be illustrated by an example: four researchers are asked to quantify the frequency of bouts of grooming, eating, travelling, and resting by a subject during a 5-min time-window, and then to assess the inter-observer agreement of their resulting data sets. Figure 2.3 represents the behaviours reported by each observer. To analyse agreement, the time-budgets produced by observers 2, 3, and 4 have been compared to observer 1, as both tallied and as sequenced data, and only kappa values have been reported (Table 2.2).

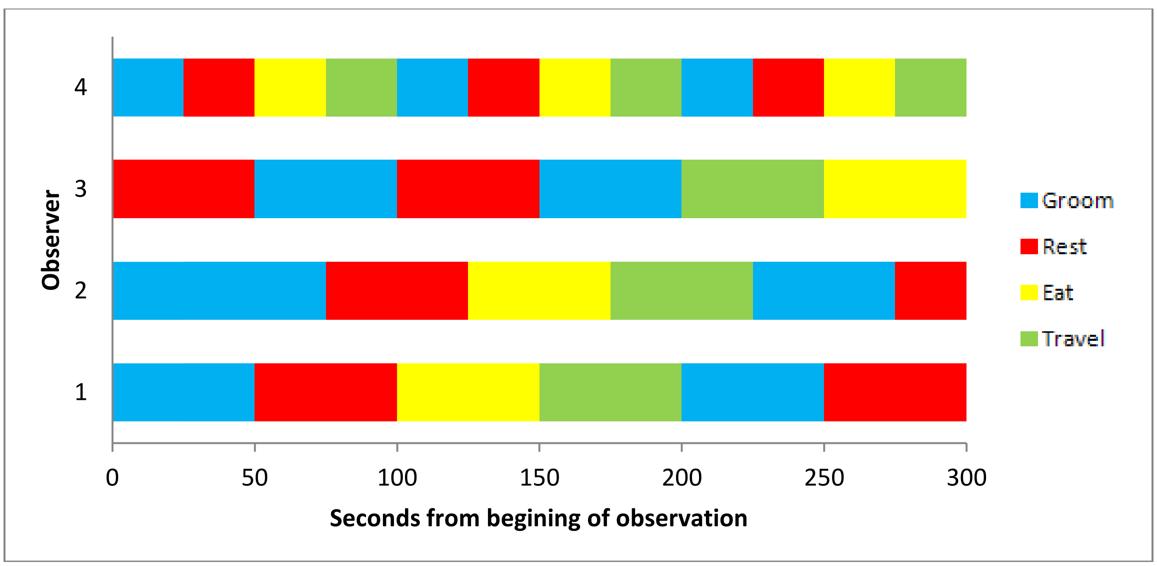

Figure 2.3 Time plots of the behaviours reported by observers 1-4 across the time-window

Table 2.2 Cohen's kappa comparing the behaviours recorded by different observers, for both durations and frequencies of both tallied and sequenced data

Data type		Observer 1 vs. 2	Observer 1 vs. 3	Observer 1 vs. 4
Duration	Tallied	0.80	1.00	0.64
	Sequenced	0.42	-0.38	0.00
Frequency	Tallied	1.00	1.00	0.43
	Sequenced	1.00	-0.09	0.22

Observer 2 produced a time-budget very similar to observer 1, except behaviours were reported 25 s later. When duration was compared as tallied data for these two observers, the agreement was almost perfect. However, when duration was compared as sequenced data, the agreement was only moderate. This is because while tallied data shows very similar totals for each behaviour, the 'blocks' of behaviours do not line up perfectly, and occasionally at different points in time the two observers disagree on what behaviour the individual was exhibiting. When frequency was compared as either tallied data or sequenced data, the agreement was perfect. The agreement for tallied behaviour was perfect because the two data sets have identical frequencies for each behaviour. Agreement for frequencies of sequenced data measure if the same behaviour happens at the same time. However, as it is very unlikely that two observers will record a behaviour at the *exact* same time, the software allows the analyst to choose a tolerance-window, and any time the same behaviour is recorded by each observer inside the tolerance-window they are considered to have agreed. In this example, the agreement for sequenced behaviour was perfect because the tolerance-window was > 25 s.

Observer 3 produced a time-budget with totals very similar to observer 1, but the behaviours were recorded in a very different order. As a result, when duration was compared as tallied data the agreement was perfect, however when it was compared as sequenced data, agreement was less than expected by chance. Similarly, when frequency was compared as tallied data the agreement was perfect, but when it was compared as sequenced data, agreement was again less than would be expected by chance. In this comparison the 'blocks' do not line up ever.

Observer 4 produced a time-budget with a very different pattern than the one produced by observer 1. When duration was compared as tallied data for these two observers, the agreement was substantial, while the agreement for sequenced data was equal to what would be expected by chance. Examination of Figure 2.3 reveals that the totals for each behaviour are relatively similar, but any overlap of the ‘blocks’ appears to be a result of chance. When frequency was compared as tallied data for these two observers, the agreement was moderate, while the agreement for sequenced data was slight. Again, examination of Figure 2.3 reveals that the frequencies of each behaviour are somewhat similar, but any time the same behaviour began within a tolerance-window appears to be a result of chance.

a: duration, tallied data

Observer 4 \ Observer 1	groom	rest	eat	travel	error	total
groom	75	-	-	-	0	75
rest	-	75	-	-	0	75
eat	-	-	50	-	25	75
travel	-	-	-	50	25	75
error	25	25	0	0	-	50
total	100	100	50	50	50	350

b: frequency, tallied data

Observer 4 \ Observer 1	groom	rest	eat	travel	error	total
groom	2	-	-	-	1	3
rest	-	2	-	-	1	3
eat	-	-	1	-	2	3
travel	-	-	-	1	2	3
error	0	0	0	0	-	0
total	2	2	1	1	6	12

c: duration, sequenced data

Observer 4 \ Observer 1	groom	rest	eat	travel	total
groom	50	0	25	0	75
rest	50	0	25	0	75
eat	0	50	0	25	75
travel	0	50	0	25	75
total	100	100	50	50	300

d: frequency, sequenced data

Observer 4 \ Observer 1	groom	rest	eat	travel	window error	total
groom	2	0	1	0	0	3
rest	2	0	1	0	0	3
eat	0	2	0	1	0	3
travel	0	2	0	1	0	3
window error	0	0	0	0	-	0
total	4	4	2	2	0	12

Figure 2.4 The confusion matrixes used to calculate the kappa for both durations and frequencies, of both tallied and sequenced data, when comparing the findings of observer 1 and observer 4. Figure 2.4a is the confusion matrix for duration of tallied behaviours, 2.4b is for the frequency of tallied behaviours, 2.4c is for the duration of sequenced behaviours, and 2.4d is for frequency of sequenced behaviours. Observer 1’s responses are represented in columns, while observer 4’s responses are along the rows. Figures 2.4a and 2.4b feature ‘-‘ in all cells except for the diagonal and the error boxes because tallied data are simply a representation of how many times the observers agree and how many times they did not. Figures 2.4c and 2.4d have all cells filled in with numbers because sequenced data represents not only when the observers agree and when they do not, but also the specifics of the disagreement.

Although the last four paragraphs hopefully explain the differences between calculating a kappa for tallied or sequenced data, the actual kappa calculations are yet to be described. Figure 2.4 shows the so-called confusion matrixes used to calculate the kappa for both durations and frequencies, of both tallied and sequenced data, when comparing the findings of observer 1 and observer 4. In the Observer software, duration and frequency of tallied data are called simply 'duration based' and 'frequency based' analysis, while duration and frequency of sequenced data are called 'duration sequence based' and 'frequency sequence based' analysis.

The confusion matrix in Figure 2.4a represents the comparison between observers 1 and 4 for the duration of tallied behaviours. The numbers inside cells on the diagonal represent the amount of time the observers agree that the subject was engaged in a certain behaviour. If one observer recorded the subject as exhibiting that behaviour for a longer amount of time, that would be reflected in the corresponding error column/row. For example, in the above example both observers agree that the subject groomed for 75 s of the observation, but observer 1 recorded that he groomed for an additional 25 s. The confusion matrix in Figure 2.4b represents the comparison between observers 1 and 4 for the frequency of tallied behaviours. This Figure is very similar in format to Figure 2.4a, except that the numbers inside the diagonal cells represent the number of times the observers agree that the subject engaged in a certain behaviour, and the number in the error column/row represents additional numbers of times a behaviour was recorded by one of the observers. For example, in the above example both observers recorded the subject groomed two times, but observer 4 recorded that he groomed an additional 1 time. In the confusion matrix in Figure 2.4c representing the comparison between observers 1 and 4 for the duration of sequenced behaviours, the numbers on the diagonal represent specific amounts of time that both observers recorded the subject as engaging in the same behaviour, and when observers disagreed on what behaviour the subject was exhibiting the amount of time was recorded in the corresponding cell. For example, in the above example there were 50 s that both observers recorded the subject as grooming, but there were also 50 s that observer 1 recorded the subject as grooming, while observer 4 recorded the subject as resting, and also 25 s that observer 4 recorded the subject as grooming, while observer 1 recorded the subject as eating. The confusion matrix in Figure 2.4d represents the comparison between observers 1 and 4 for the frequency of sequenced behaviours. This format is very similar to Figure 2.4c, except that the numbers inside the diagonal cells represent the frequency with which the observers agree that the subject engaged in a certain behaviour at a specific time, within a set out tolerance window (e.g. a 30 s grace period). When observers disagreed on what behaviour the subject was exhibiting at a certain time, it was recorded in the corresponding cell. For example, in the above example the observers agreed that the subject groomed 2 times, but there were also 2 times that observer 1 recorded the subject as grooming, while observer 4 recorded the subject as resting, and also 1 time that observer 4 recorded the subject as grooming, while observer 1 recorded the subject as eating. Additionally, if the observers agreed upon what behaviour the subject was exhibiting, but the start time was outside of the tolerance window, this is registered in the 'window error' column/row. There were no instances of this in this example. Using each of these matrices, kappa is then calculated using the formula in Figure 2.1, where the proportion of observed agreement equals the sum of the numbers on the top left to bottom right diagonal inside the box divided by the total sum of the numbers inside the box, and the proportion of agreement due to chance equals the total of the first row multiplied by the total of the first column divided by the square of the total sum of the numbers inside the box, added to this same calculation for each subsequent column/row. For example, the kappa calculation for Figure 2.4a can be found in Figure 2.5, and agrees with the value presented in Table 2.2.

$$\textbf{Proportion of observed agreement} = \frac{\mathbf{75 + 75 + 50 + 50}}{\mathbf{350}} = \mathbf{0.71}$$

$$\textbf{Proportion agreement expected due to chance}$$

$$= \left(\frac{\mathbf{100 \times 75}}{\mathbf{350^2}}\right) + \left(\frac{\mathbf{100 \times 75}}{\mathbf{350^2}}\right) + \left(\frac{\mathbf{50 \times 75}}{\mathbf{350^2}}\right) + \left(\frac{\mathbf{50 \times 75}}{\mathbf{350^2}}\right) = \mathbf{0.18}$$

$$\textbf{Kappa} = \frac{\textbf{Proportion observed agreement} - \textbf{Proportion agreement expected due to chance}}{\mathbf{1} - \textbf{Proportion agreement due to chance}}$$

$$\textbf{Kappa} = \frac{\mathbf{0.71 - 0.18}}{\mathbf{1 - 0.18}} = \mathbf{0.64}$$

Figure 2.5 Kappa calculations for the comparison between observers 1 and 4 for the duration of tallied behaviours

Before computerised methods of data collection became practical, data were almost always expressed as tallied samples. However, increasingly more and more data are being collected as sequenced samples. This is because sequenced data are no more difficult to collect with these computer-aided methods than tallied data, but contain far more information, as the sequence of behaviours is available as well. Furthermore, once recorded, sequenced data can easily be converted to tallied data, but the reverse is not possible. However, often the research question that is being asked does not require the detail provided by sequenced data, and tallied data would have been sufficient. Jansen et al. (2003) argue that the way the data will be used should dictate the application of the agreement assessment, not the collection type. They state that agreement assessments should be conducted on sequenced data when the sequence of events is important to the research question, and on tallied data when only overall totals are important to the research question (Jansen et al., 2003). However, in the example above, observers 1 and 3 are in complete agreement in tallied data, but worse than chance agreement in sequenced data. Even if a research question only requires results based on tallied data, inspection of the time plots reflective of these two data sets should clearly indicate that there is a problem somewhere in the data collection process (perhaps the two observers have not been trained properly). Instead it can be suggested that sequenced analysis of agreement is worthwhile whenever possible if inter-observer agreement is in question.

2.7 Conclusions

In this thesis, all assessments of agreement and reliability have been conducted in accordance with what is presented in Table 2.1, however in some instances additional assessments have been presented to facilitate cross-study comparison. Whenever assessment of inter-observer agreement is conducted with Observer software, all results are based on sequence based analyses.

2.8 References

Barnhart, H.X., Haber, M., Song, J., 2002. Overall concordance correlation coefficient for evaluating agreement among multiple observers. Biometrics 58, 1020-1027.

Bland, M.J., Altman, D.G., 1986. Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet 327, 307-310.

Carrasco, J.L., Jove, L., 2003. Estimating the Generalized Concordance Correlation Coefficient through Variance Components. Biometrics 59, 849-858.

Chen, C., Barnhart, H.X., 2008. Comparison of ICC and CCC for assessing agreement for data without and with replications. Comput. Stat. Data Anal. 53, 554-564.

Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences, 2nd Ed. Lawrence Erlbaum Associates, New Jersey.

Cooper, J.O., Heron, T.E., Heward, W.L., 2007. Applied Behavior Science, 2nd Ed. Pearson, Upper Saddle River, N.J.

de Vet, H., 2005. Observer Reliability and Agreement. Encyclopedia of Biostatistics , 1-5.

de Vet, H.,C.W., Terwee, C.B., Knol, D.L., Bouter, L.M., 2006. When to use agreement versus reliability measures. J. Clin. Epidemiol. 59, 1033-1039.

Dohoo, I.R., Stryhn, H., Martin, S.W., 2009. Veterinary Epidemiologic Research, 2nd ed. VER, Inc., Charlottetown, P.E.

Ferguson, C.J., 2009. An effect size primer: A guide for clinicians and researchers. Professional Psychology: Research and Practice 40, 532-538.

Gibbons, J., Lawrence, A., Haskell, M., 2009. Responsiveness of dairy cows to human approach and novel stimuli. Appl. Anim. Behav. Sci. 116, 163-173.

Gisev, N., Bell, J.S., Chen, T.F., 2013. Interrater agreement and interrater reliability: Key concepts, approaches, and applications. Research in Social and Administrative Pharmacy 9, 330-338.

Jansen, R., Wiertz, L., Meyer, E., Noldus, L., 2003. Reliability analysis of observational data: Problems, solutions, and software implementation. Behavior Research Methods 35, 391-399.

Kottner, J., Audigé, L., Brorson, S., Donner, A., Gajewski, B.J., Hróbjartsson, A., Roberts, C., Shoukri, M., Streiner, D.L., 2011. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J. Clin. Epidemiol. 64, 96-106.

Landis, J.R., Koch, G.G., 1977. The Measurement of Observer Agreement for Categorical Data. Biometrics 33, pp. 159-174.

Lin, L.I., 1989. A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometrics , 255.

Mitchell, S.K., 1979. Interobserver agreement, reliability, and generalizability of data collected in observational studies. Psychol. Bull. 86, 376-390.

Noldus Information Technology, 2003. The Observer: Professional System for Collection, Analysis, Presentation and Management of Observational Data. Reference Manual. Version 5.0. Tracksys Ltd., Nottingham.

Roberts, B.W., DelVecchio, W.F., 2000. The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. Psychol. Bull. 126, 3-25.

Sheskin, D.J., 2003. Handbook of Parametric and Nonparametric Statistical Procedures. Chapman and Hall/CRC.

Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlations: Uses in assessing rater reliability. Psychol. Bull. 86, 420-428.

Sprinthall, R.C., Fisk, S.T., 2002. Basic Statistical Analysis. Prentice Hall, Englewood Cliffs, NJ.

Tinsley, H.E., Weiss, D.J., 2000. Interrater reliability and agreement, in: Tinsley, H.E., Brown, S.D. (Eds.). Academic Press, San Diego, California, pp. 95-124.

Weinberg, S.L., Abramowitz, S.K., 2002. Data Analysis for the Behavioral Sciences using SPSS. Cambridge University Press, Cambridge.

3 QUANTITATIVE AND QUALITATIVE BEHAVIOURAL AND FAECAL GLUCOCORTICOID RESPONSES OF CATS TO A SINGLY HOUSED ENVIRONMENT

3.1 Abstract

The behaviour and faecal glucocorticoid metabolite (FGM) responses to caging were examined in six shelter cats caged for 30 days. Continuous focal observations of the activity, location in the cage, and posture were conducted from video recordings for one 24-h period/week/cat. Cat-Stress-Scores (CSS) were recorded daily. Faecal samples were collected for analysis of FGM. The percentage time spent eating increased, while percentage time spent grooming decreased, from week 1 to week 2. CSS declined significantly from week 1 to week 2. Log_e FGM concentrations were greater in week 1 than in week 5. A post-habituation time budget of the behaviour of the cats in the cages showed that the cats were located on the shelf almost half of the time – tending towards more than the expected value. This suggests that it may have been a resource of value to the cats, and that its inclusion in enclosure design may be important. Quantitative and qualitative behavioural data indicated that there was an initial stress response to caging that stabilised after the first week, while FGM concentration took longer to stabilise.

3.2 Introduction

Every year, large numbers of stray, abandoned, or feral cats and kittens end up in shelters (Rochlitz, 2000). In Canada, while 45% of cats who enter shelters are adopted, 46% are euthanized (the remaining 9% were returned to their owner, transferred to another organisation, or died in the shelter) (Canadian Federation of Humane Societies, 2010). Negative behaviour has been cited as the number-one reason a particular cat is euthanized (Wenstrup and Dowidchuk, 1999) while positive behaviour (such as social greeting) has been cited as the number-one reason an adopter chooses a particular cat (Weiss et al., 2012). Gourkow and Fraser (2006) have suggested that the problem behaviours repelling adopters may be a product of fear and stress experienced by cats as a result of the shelter environment. Shelters are potentially stressful environments for cats for a number of reasons, including strict spatial limitations, altered routine, exposure to unfamiliar stimuli, absence of familiar stimuli, social deprivation and lack of physical complexity (Carlstead et al., 1993). This stress could also be linked to illness, such as the increased likelihood of developing an upper respiratory tract infection (Tanaka et al., 2012)

When an individual encounters a stressor, activation of the hypothalamic-pituitary-adrenal (HPA) axis can occur (Moberg, 2000). The use of a non-invasive method of measuring glucocorticoid production and metabolism – such as faecal glucocorticoid metabolites – permits the collection of samples in a way that is less disruptive to the individuals, thereby removing the confounding effects of stress caused by blood sampling (Hodges and Heistermann, 2003; Whitten and Brockman, 1998) and provides a daily integrated response to a longer-term stressor, such as housing. A number of ways have been used to attempt to quantify the behavioural responses to stress in cats. The Cat-Stress-Score (CSS) is an ordinal rating system developed by Kessler and Turner (1997) which considers 11 behavioural/postural categories, and assigns scores from 1-7 based on provided definitions that interpret the severity of the response, and is often used to assess stress in shelter cats (e.g. Dybdall et al., 2007; Hawkins et al., 2004). Additionally, it has been suggested that when stressed, cats rarely actively exhibit abnormal behaviour and it is more likely that they will just exhibit normal behaviours at abnormal frequencies (Rochlitz, 1999). For this reason, continuous behavioural observation and analysis of differences between individuals and analysis over time have the potential to reveal subtle behavioural changes that may be reflective of stress.

In a shelter setting, staff often have little time to dedicate to each individual animal, and behavioural signs of stress can be helpful when identifying individual cats that find the shelter environment particularly stressful.

Two of the reasons that have been cited for why shelters are stressful are: 1) exposure to unfamiliar stimuli, and 2) the absence of familiar stimuli. An individual is likely to experience the stress associated with these factors most acutely immediately following arrival at the shelter, and thus their impact should reduce with habituation. Studies using the CSS provide some evidence for this phenomenon (Kessler and Turner, 1997), but the CSS response has yet to be studied alongside other methods, such as faecal glucocorticoid metabolite (FGM) analysis and continuous quantitative behavioural observation that might be able to provide confirmatory evidence of this habitation. Furthermore, to my knowledge, no study has published a post-habituation time budget for singly housed domestic cats in cages. This information would be useful in not only identifying abnormal behaviour, but data describing how the cat is utilising the space may provide information as to what aspects of the enclosure are preferred by the cats, helping to guide future cage design.

This study aimed to: 1) investigate changes in physiological and behaviour coinciding with habituation to the environment, and 2) provide a post-habituation time budget for the behaviour exhibited by cats in singly caged housing.

3.3 Methods

3.3.1 Subjects

This study was conducted using six domestic cats obtained from a local animal shelter, of which three were intact males and three were females of unknown reproductive status, all of which tested negative for feline immunodeficiency virus and feline leukemia virus. The mean length of stay in the shelter before transport approximately 5 km to the study site was 2.6 ± 1.92 days (±SD), and mean estimated age of cats included in this study was 2.5 ± 1.05 years. Two of the cats had been surrendered by their owner and four of them were strays. Three additional cats were present at various points during the seven-weeks of study, but were removed due to inappetence or upper respiratory tract infection. At any one time, there were between 3 and 6 cats in the study.

3.3.2 Housing and management

On arrival, each cat was placed into a stainless steel cage (58 × 79 × 79 cm) in a dedicated research room. The cages were in two rows of three cages approximately 1.5 m apart which faced each other so that all cats could see, hear, and smell each other. Cats were provided with 12 h of fluorescent lighting and 12 h of darkness with infrared illuminated lighting. The fluorescent lights came on at 06:00 h and went off at 18:00 h each day. Both banks of cages had three black-and-white CCTV cameras (Panasonic, Germany) and two infrared illuminated microlights (880 Infra Red Illuminator, Dennard, UK) directed towards the opposite cages so that each cage had one dedicated camera, and the light was positioned so that each cage was sufficiently illuminated to permit behavioural observation. Within each cage, there was a shelf (55 × 20 cm) elevated 32 cm off the floor covered in cardboard, a pink hand towel on the shelf, a litter pan (33 × 28 × 13 cm), a small plastic ball with a bell inside, and food and water dishes secured to the cage door. Of the places for the cat to be located, the shelf represented 20%, the litter pan 16%, and the floor 64% of the available space. The cats were free to shift any movable objects throughout the day, but twice a day when cages were cleaned the items were rearranged.

Cats were offered 50 g of dry cat food (Adult Indoor, Nutrience by Hagan) twice daily: once at 09:30 h and once at 16:30 h. Litter pans were cleaned, and water bowls were filled by animal care technicians twice daily: once between 08:00 h and 09:00 h, and once between 15:00 h and 16:00 h. The room was maintained at a temperature of 20^{o}C (range ± 2). Researchers and animal care staff interacted with the cats only as much as was necessary in the routine feeding, cleaning, and experimental protocol. A concurrent study investigating temperament was conducted using the same cats, which involved removing and socialising with the cats for half an hour once per week.

3.3.3 Continuous quantitative behavioural observations

Cats were time-lapse recorded (VCRs: CTR-3024, Computar, UK; multiplexer: Sprite dx, Dedicated Micros, UK) 24 h a day for 30 days. One 24 h period per week was analysed for each cat's general activity, location within the cage, and posture using continuous observation (Martin and Bateson, 2007). Recordings were filmed at 50 frames/s, and viewed at 10 times the actual speed. The weekly sample periods were every seven days beginning immediately following arrival at the facility. The ethogram for location, posture and activity is given in Table 3.1. A behavioural bout was considered to have concluded if the animal ceased to perform the behaviour for at least 5 s. The analysis of behaviour was conducted by two different observers, using Noldus Observer 5 (Noldus Information Technology, Wageningen, The Netherlands). Observations were evenly distributed by both cat ID and week in study between observers.

3.3.4 CSS recordings

Once daily (at 16:00 h), two researchers simultaneously stood in front of cages for 10 min, to habituate the cats to their presence, and then assigned the cats a score from 1 (fully relaxed) to 7 (terrorised) in 11 behavioural/postural categories (Kessler and Turner, 1997). Researchers were blinded to each other's scoring, and assessments were repeated within 15 min.

3.3.5 FGM measurement

Faecal collection

When litter pans were cleaned all faecal samples present were homogenised, and placed into 30 ml Nalgene tubes. The samples were then temporarily stored at -4^{o}C and then stored at -20^{o}C until extraction. A mean of 30 ± 8.45 samples were collected from each cat.

Hormone extraction and analysis

Samples were extracted and analysed according to Möstl and Palme's (2008) protocol using EIA 9.3 11-oxoaetiocholanolone (Lab-code: 72-alt; EIA first described: Palme and Möstl (1997) and previously validated for use in cats (Palme et al., 2001). Samples were run sequentially in the order they were produced. In the current study, inter- and intra-assay coefficients of variation were 8.9% (n=11) and 3.5% (n=385).

Table 3.1 Ethogram for quantitative behavioral observations (adapted from UK Cat Behaviour Working Group, 1995)

Behavioral Category	Behavior	Definition
Activity	Resting	Cat remained generally inactive
	Grooming	Cat licked its body or licks its paw and passed the paw over its head
	Eating	Cat consumed (or appeared to consume) food
	Drinking	Cat lapped water
	Manipulation	Cat manoeuvred or attempted to manoeuvre an object with its paw
	Locomotion	Cat moved position within its enclosure
	Out of cage	Cat was not present in cage
	Other activity	Any activity not defined above
Location	On shelf	Cat was positioned on top of the provided shelf
	In litter pan	Cat was positioned in the provided litter pan
	Behind litter pan	≥ 50% of the cat was positioned behind the provided litter pan
	Cage front	Cat was positioned at the cage bars at the front of the enclosure
	Floor other	Cat was positioned on the cage floor, other than at the cage front or behind the litter pan
	Out of cage	Cat was not present in cage
	Other location	Any location not defined above
Posture	Lying head down	One side of cat was in complete contact with the ground, head on side or extended
	Lying head up	One side of cat was in complete contact with the ground, head not in contact with ground
	Sitting	Pads of the front paws were on the ground with the front legs straight and the rump on the ground
	Standing	Cat was positioned with four paws on the ground, rump raised
	Locomotion	Cat moved position within its enclosure
	Out of cage	Cat was not present in cage
	Other posture	Any posture not defined above
View obscured	View obscured	Sight lines to the cat were obstructed

3.3.6 Statistical analyses

Continuous quantitative behavioural observation.

The inter-observer agreement of the two observers was evaluated for both durations and frequencies of behaviours, from six 24-h periods (one of each cat during the first week) using methods of percentage of agreement, Pearson's *r*, and Cohen's kappa test statistic (Jansen et al., 2003). In the interest of cross-study comparison, all have been calculated. Minimum acceptability of results was set at ≥80% for percentage agreement (Cooper et al., 2007), ≥0.8 (or a strong effect) for Pearson's r (Ferguson, 2009), and >0.6 (meaning substantial or better) for Cohen's kappa statistic (Landis and Koch, 1977).

For each behavioural class, percentage of time spent performing each of the specific behaviours was calculated for each 24-h period analysed. Two-way ANOVA models were used to investigate the effect of cat ID and week on each outcome variable, and Bonferroni adjustments for assessing multiple behaviours were carried out within each behavioural class. Where necessary to satisfy model assumptions, data were transformed by the arcsine square root function. When model assumptions could not be met, two Friedman's tests were carried out for that variable, switching cat ID and week between treatment and block, in order to generate test statistics for each variable. To assess the impact of within-cat correlation structure, repeated measures ANOVAs were run and generally, no changes in the result were found. Furthermore, for the two outcome variables with a significant effect of week, linear mixed models showed no evidence of autocorrelation within cats. For variables where week was found to be a significant factor, Bonferroni pairwise comparisons were used to compare weeks. All analyses were conducted using Minitab® 15 statistical software.

CSS recordings

Daily CSSs were compared between researchers with a Wilcoxon signed rank test. Weekly median values of the two researchers' scores were used for further analyses. Friedman's test was used to analyse median weekly CSS across weeks, with week in study blocked by cat ID. Sign tests were used to analyse CSS differences between specific weeks.

FGM measurement

If more than one faecal sample was produced by a cat in one day, an average of the FGM values was used (Touma and Palme, 2005). FGM concentration values were $\log_e$ transformed and linear models were used to investigate the effects of cat ID and week. Bonferroni pairwise comparisons were used to compare weeks.

Time budget

The percentages of time that each cat spent performing each of the specific behaviours in all three behavioural classes were calculated for the four 24-h observations conducted on weeks 2-5. As some behaviours were found to vary significantly (see below) between week one and all other weeks, data from week 1 was not included in the time budget. One sample t-tests were used to compare time spent on each resting surface (cage floor, litter pan, and shelf) to the expected distribution based on available space.

3.3.7 Ethical approval

This project was approved by UPEI's Animal Care Committee under protocol number 09-051, and followed the guidelines of the Canadian Council of Animal Care's "Guide to the Care and Use of Experimental Animals".

3.4 Results

3.4.1 Continuous quantitative behavioural observation

The inter-observer agreement of the two observers was found to be higher for durations than for frequencies in all three measures (Table 3.2). Since values for durations satisfied all of the minimum acceptability standards set out in the methods, while the values for frequencies failed to meet the standards for two out of three of the measures, all subsequent analyses were only conducted on the durations of the behaviours.

Table 3.2 Mean (±SD) inter-observer agreement between 2 observers of six 24-h observations

	Percentage agreement	Pearson's *r*	Cohen's kappa
Duration/sequence based	86 (4.8)	1.00 (0.01)	0.82 (0.06)
Frequency/sequence based	52 (9.0)	0.92 (0.06)	0.48 (0.08)

Cat ID was found to be a significant factor for a number of behaviours: in the behavioural class activity, cat ID was significant for drinking and grooming; in the behavioural class location, cat ID was significant for 'in litter pan' and 'on shelf'; in the behavioural class posture, cat ID was significant for lying head down, sitting, and standing. Week was also a significant factor for eating and grooming (Table 3.3).

The percentage of time per day spent eating during week 1 was significantly less than during weeks 2 to 5 (P<0.05), but after week 1 there were no significant differences between weeks. The percentage of time per day spent grooming during week 1 was significantly greater than during weeks 2 to 5 (P<0.001), but after week 1 there were no further significant differences between weeks (Table 3.4).

Table 3.3 Effects of 'cat' and 'week in study' on the percentage of time spent per 24-h period engaged in each behavior

Behavioral Class	Behavior	Test	Cat ID		Week	
			Test statistic	P value	Test statistic	P value
Activity	Drinking	2-way ANOVA	$F_{5,20}$=9.08	0.004	$F_{4,20}$=4.25	0.084
	Eating	2-way ANOVA	$F_{5,20}$=2.45	0.490	$F_{4,20}$=8.11	0.004
	Grooming	2-way ANOVA	$F_{5,20}$=5.77	0.014	$F_{4,20}$=10.66	0.004
	Manipulation[a]	2-way ANOVA	$F_{5,20}$=2.41	0.511	$F_{4,20}$=1.56	1.000
	Locomotion[a]	2-way ANOVA	$F_{5,20}$=4.38	0.049	$F_{4,20}$=0.77	1.000
	Other	Friedman's Test	S_5=5.80	1.000	S_4=4.27	1.000
	Resting	2-way ANOVA	$F_{5,20}$=2.24	0.630	$F_{4,20}$=0.23	1.000
Location	Behind litter pan	Friedman's Test	S_5=8.82	0.720	S_4=3.79	1.000
	Cage front	Friedman's Test	S_5=8.87	0.714	S_4=7.07	0.792
	Floor other	Friedman's Test	S_5=10.03	0.444	S_4=2.40	1.000
	In litter pan	Friedman's Test	S_5=17.00	0.024	S_4=7.87	0.582
	On shelf	2-way ANOVA	$F_{5,20}$=5.37	0.018	$F_{4,20}$=0.62	1.000
	Other	Friedman's Test	S_5=5.00	1.000	S_5=8.00	0.552
Posture	Locomotion[a]	2-way ANOVA	$F_{5,20}$=4.38	0.042	$F_{4,20}$=0.77	1.000
	Lying head down	2-way ANOVA	$F_{5,20}$=5.07	0.024	$F_{4,20}$=0.79	1.000
	Lying head up	2-way ANOVA	$F_{5,20}$=4.25	0.048	$F_{4,20}$=0.90	1.000
	Other	Friedman's Test	S_5=5.00	1.000	S_4=8.00	0.552
	Sitting	2-way ANOVA	$F_{5,20}$=13.16	0.004	$F_{4,20}$=2.03	0.774
	Standing	2-way ANOVA	$F_{5,20}$=9.36	0.004	$F_{4,20}$=1.40	1.000
View Obscured	View Obscured	Friedman's Test	S_5=5.00	0.416	S_4=4.00	0.406

Note. P-values Bonferroni corrected for testing on total of 7 outcome variables for Activity, 6 outcome variables for Location, and 6 outcome variables for Posture.
Test statistic: $F_{df1,df2}$ statistic for ANOVA, S_{df} statistic for Friedman's test
[a]Transformed using arcsine square root

Figure 3.1 shows the distribution of the behaviours eating and grooming across the day. Duration of all instances of these behaviours exhibited during the study period was summed within the hour the instance was initiated. These sums were then plotted against the hours of the day. Visual inspection of the plot was used to identify windows of peak exhibition of these behaviours for use in future analyses geared toward targeting these potential behavioral indicators of stress. The hour in which most grooming was initiated was at 02:00 h, with minor peaks at 11:00 h and 23:00 h. The hour in which most eating was initiated was at 09:00 h, followed by 15:00/16:00 h, corresponding closely to the feeding times. Mean bout lengths were also calculated for both behaviours. The mean duration of grooming bouts was 251±74.9 s (or, 4 min 11 s), and the mean duration of eating bouts was 300 ±113.5 s (or, 5 min).

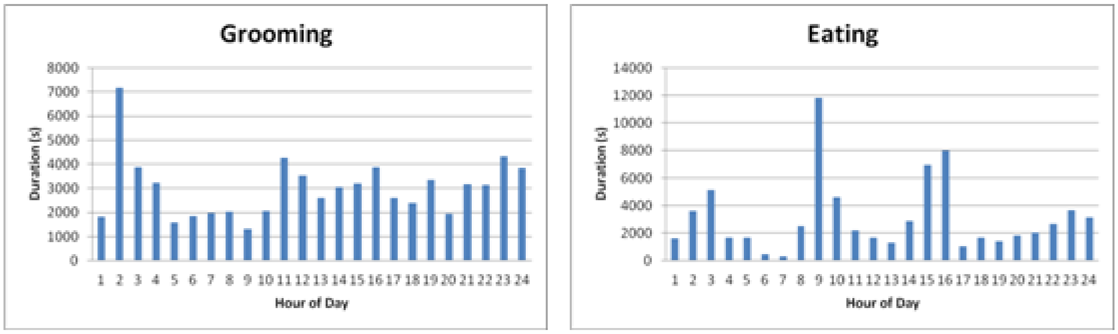

Figure 3.1 Durations of grooming and eating within each hour of the day that the behaviour was initiated, summed over the five weekly 24-h observation periods.

3.4.2 CSS recordings

A one-sample Wilcoxon signed rank test showed no significant difference ($P=0.794$) between daily CSS recorded by researchers, thus the median of the researchers' daily scores was calculated to create one daily CSS value for each cat. Weekly median values of these daily scores were used for further analyses. Friedman's test (week blocked by cat) revealed a significant difference in CSS across weeks ($S= 16.31$, d.f.=4, $P=0.003$). Sign tests revealed that the CSS for week 1 was significantly greater than during weeks 2 to 5 ($P<0.05$), but after week 1 there were no significant differences between weeks (Table 3.4).

3.4.3 FGM measurement

There was a significant effect of week on the $\log_e$ FGM concentration ($F_{4,20}=4.21$, $P=0.012$). Week 5 was significantly lower than week 1 ($P<0.05$). There were no other significant differences between weeks (Table 3.4).

Table 3.4 Weekly summary statistics for percentages of time spent per 24-h period eating and grooming, CSS, and FGM (n=6)

	Eating Mean (±SD)	Grooming Mean (±SD)	CSS Median (IQR)	Log $_e$ FGM (ng steroid/g faeces) Mean (±SD)
Week 1	1.4^{a} (0.94)	5.4^{a} (2.26)	2.4^{a} (0.88)	5.4^{a} (1.07)
Week 2	2.8^{b} (0.55)	2.3^{b} (0.89)	2.1^{b} (0.23)	4.5^{ab} (1.01)
Week 3	3.6^{b} (1.34)	2.2^{b} (1.43)	2.0^{b} (0.12)	4.4^{ab} (0.73)
Week 4	3.5^{b} (0.80)	2.6^{b} (0.80)	2.0^{b} (0.07)	4.5^{ab} (0.65)
Week 5	3.4^{b} (0.61)	2.1^{b} (1.40)	2.1^{b} (0.27)	4.1^{b} (1.03)

^{ab} Within a column, different superscripts indicate a significant difference between weeks (P<0.05).

Table 3.5 Percentages of time spent per 24-h period engaged in each behavior, within each behavioral class, in weeks 2-5, n=6

Behavioral Class	**Behavior**	**Median (IQR)**
Activity	Drinking	1 (1.0)
	Eating	3 (1.2)
	Grooming	2 (2.0)
	Manipulation	1 (0.8)
	Locomotion	2 (0.8)
	Other	0 (0.1)
	Resting	91 (2.4)
Location	Behind litter pan	0 (11.8)
	Cage front	11 (6.1)
	Floor other	23 (39.5)
	In litter pan	1 (0.3)
	On shelf	53 (50.3)
	Other	0 (0.0)
Posture	Locomotion	2 (0.8)
	Lying head down	54 (14.3)
	Lying head up	29 (14.6)
	Other	0 (0.0)
	Sitting	11 (8.5)
	Standing	2 (2.3)

Note. For each cat, the median of each weekly 24-h value over weeks 2-5 was calculated and then the median and IQR for each behavior, over weeks 2-5, for the 6 cats was calculated. As not all behaviours had normal distributions, medians were presented for all to be consistent.

3.4.4 Time budget

As stated above, week 1 data were excluded from the analysis of the time budget for habituated cats because some significant differences were found between week 1 and all other weeks. The following time budgets are for weeks 2-5. In the behavioural class 'activity', cats spent the largest percentage of their time 'resting' (median=91, IQR=2.4, n=6). In the behavioural class 'location', cats spent the largest percentage of their time 'on shelf' (median=53, IQR=50.3, n=6). In the behavioural class 'posture', cats spent the largest percentage of their time 'lying head down' (median=54, IQR=14.3, n=6). Table 3.5 shows the median percentages of time per 24-h period engaged in each behaviour, within each behavioural class, in weeks 2-5. One sample t-tests revealed that cats were not located on the floor significantly differently from expected (t_5=0.83, P=0.446), were located in the litter pan significantly less than expected (t_5=5.21, P=0.003), and were located on the shelf close to significantly more than expected (t_5=-2.22, P=0.077).

3.5 Discussion

This study presents some novel and interesting findings on the habituation of singly housed cats to a cage, and on their behaviour post-habituation.

3.5.1 Continuous quantitative behavioural observations

Analysis of the continuous behavioural observations revealed that cat ID was a significant factor for many behaviours. This is not surprising, due to the well-developed concept of individuality in the domestic cat. This concept is common not only in popular discourse and literature, but also in academic research (Mendl and Harcourt, 2000). There is the potential for many factors to contribute to a cat's individuality, from more obvious distinctions such as age and sex, to concepts that are more difficult to define and/or study such as genetic differences, behavioural styles, and experience. The limited sample size in the current study may have reduced the power to identify subtle differences between certain weeks, but some larger differences between weeks were observed. As there was variation within the subjects in terms of cat-level characteristics such as sex, age, and size, it is possible that these differences may have influenced their response to the initial shelter conditions and then cage confinement during the study. Notably, Dybdall et al. (2007) found that surrendered cats experienced significantly greater behavioural stress after entering shelter conditions than stray cats. As both types of cats were included in the current study, these differences could have potentially influenced the results. Unfortunately, due to the small sample size, analysis of this and other cat-level characteristics was omitted.

Percentage of time spent grooming was found to decline significantly from week 1 to week 2, which both supports and conflicts with this finding in the literature. Iki et al. (2011) found that grooming increased after exposure to a stressor. Conversely, Griffith et al. (2000) found that cats exposed to a feline facial pheromone thought to have *anxiolytic* properties exhibited an *increase* in grooming. Although this pheromone has been used to treat behaviour problems associated with anxiety (Frank et al., 1999), little is known about its function or mechanism of action (Pageat and Gaultier, 2003) or the validity of its anxiolytic effect. Nevertheless, support for increased grooming being a behavioural sign of stress can also be found in other species, and in primates it has been validated as a behavioural indicator of anxiety through the administration of anxiolytic and anxiogenic drugs, possibly acting as a displacement behaviour (Maestripieri et al., 1992). The results of this study, when considered in the context of previous studies, suggest that the direction of the relationship between grooming and stress is not particularly straightforward, and might not be useful for identifying stressed individuals in a shelter setting.

The increased percentage of time spent eating from week 1 to week 2 found in the current study could be interpreted as a sign of diminished stress. Tanaka et al. (2012) found that for cats newly introduced to a shelter, food intake and CSS were negatively correlated, and that cats with high stress scores were 5.6 times more likely to develop an upper respiratory tract infection. Griffith et al. (2000) found that cats exposed to a pheromone thought to have anxiolytic effects exhibited a significantly higher frequency of interest in food, and of eating than did cats not exposed to the pheromone. Furthermore, Smith et al. (1994) found latency to consume food after presentation reduced significantly over time. Conversely, Gouveia et al. (2011) found eating (coupled with drinking) was negatively related to time in the shelter. However, it is worth noting that Gouveia et al. were investigating the long-term stressor of enduring confinement to a shelter (7+ years) while the current study was investigating the short-term stressor of habituation to a shelter (30 days) and thus, processes other than habituation may be reflected in the results.

Visual inspection of the distribution of grooming and eating revealed that these behaviours were exhibited differentially throughout the day. If observational analysis of behaviour is not possible or impractical from a full 24 h period, it may be helpful to identify windows of time when these behaviours are likely to be exhibited. In order to keep up with the daily footage being recorded, it is logical to analyse only as much as can be observed in one day. In this study, the researcher found that 8 h of time-lapsed footage was the maximum that could be analysed per day before fatigue set in. This equates to one third of the day, which is a large enough proportion to assume it would be a relatively representative sample of the day. This 8 h could be consecutive footage, or broken in to smaller windows throughout the day. However, making windows too small has the potential to falsely reduce mean bout lengths by failing to capture the full duration of a bout. The greatest amount of grooming was initiated at 02:00 h, and another two potential peaks were discernible at 11:00 h and 23:00 h. The greatest amount of eating was initiated at 09:00 h, with a smaller peak at 15:00/16:00 h. If the 8 h footage sample was broken down into two 4 h windows, the periods of 23:00-03:00 h and 07:00-11:00 h would capture much of the grooming and eating respectively, while also representing behaviours exhibited during both light and dark periods. Plotting sums of durations based on the hour in which the behaviour was initiated is not as ideal as plotting amount of time engaged in the behaviour per hour. However, as this analysis was not central to the research question of this study, and restructuring the data output produced by Observer software to calculate amount of time engaged in each behaviour per hour would be time consuming, the analysis produced here was deemed sufficient to visualise the distribution of the exhibition of these behaviours across the day. The agreement between when behaviours were initiated and the actual amount of time engaged in each behaviour per hour would be biased by bout duration. If bouts were long and were initiated near the end of an hour, the majority of the behaviour would actually be exhibited in the following hour or hours. However if the bout was short it is more likely that the majority of the behaviour would be exhibited during the hour in which it was initiated. As the mean bout length of both grooming and eating was between 4 and 5 min, this bias is unlikely to have affected the distribution to a large degree.

There were a number of behaviours identified as potential indicators of stress in previous studies that were not identified in the current study: most notably hiding behaviour (Carlstead et al., 1993; Rochlitz, 1999; Smith et al., 1994). This could be due to differences in the methods of recording, amounts of data, and/or observer interaction. First, the varying methods of recording the behaviour (i.e. frequency vs. duration and percentage of time spent) may have identified different nuances of the behaviour. Second, while Gouveia et al. (2011) provided data on both frequency and duration, their total sampling time per cat was only 15 min. Carlstead et al. (1993) however, amassed an impressive 744 h per cat. Finally, studies such as Gouveia et al.'s (2011) where the observer was in the pen with the cats introduces the potential for observer presence to influence behaviour, while studies such as Carlstead et al. (1993) which recorded behaviour from video tape do not suffer from such issues. However, sometimes detailed behaviours can be difficult to analyse from video tape. Clearly methodological issues other than type of data collected have the potential to influence results. Unfortunately, the current study lacked high inter-observer agreement in frequency data, and thus only data on duration and percentage of time spent could be analysed, so direct comparison with studies using frequency data was not possible. There is also the potential influence of cage contents. Large numbers of objects behind which cats can hide may result in more or longer instances of hiding. However, it is likely that if a cat is motivated to hide by fear or stress, they will find a way to do so regardless of the minimal cage contents. It is also possible that there were differences in behavioural definitions. In the current study, the location 'behind litter box' (easily interpreted as hiding) was defined as ≥ 50% of the cat positioned behind the provided litter pan. Other studies have been less clear about exactly how they defined the cat's locations within the cage. Also the current study investigated the stressor of being introduced to cages in research, while the other studies examined either introduction to an actual busy shelter environment, or exposing cats to a specific stressor under experimental conditions. It is possible that the conditions of the current study were not sufficiently stressful to evoke the behavioural response observed in other studies. Finally, perhaps the changes in behaviours in the current study were not signs of diminishing stress associated with habituation, but were of boredom associated with exposure to an unchanging environment. It is possible that increased eating could be a sign of boredom (Macht, 2008); however, it is just as conceivable that a cat might limit food intake in response to change, and only return to regular rates when the cat is more accustomed to its environment. Additionally, as grooming is thought to be associated with boredom in many species (Humans – Daly et al., 1983; Dairy cows - DeVries et al., 2007); Vervet monkeys - Fruteau et al., 2009), the fact that grooming was found to *decrease* from week 1 to week 2 and then remain stable, suggests this variable was not evidence of growing boredom, but rather evidence of initial stress due to caging, and subsequent habituation to the environment.

3.5.2 Cat-Stress-Scores

The CSS in the current study supported the findings of Kessler and Turner (1997) that the CSS reduces significantly in the first week, and tends to more-or-less stabilise after that, likely due to habituation. Interestingly, the quantitative measurement of the times spent eating and grooming also changed after the first week. Perhaps this is the amount of time it takes for cats to habituate to a novel caged environment. Smith et al. (1994) suggests that most behaviours stabilise after the first four days in a shelter, with smaller changes occurring within the first month. The main difference between the findings in the current study and those of Kessler and Turner (1997) was one of magnitude. In the present study, only one of the cats had a median score above 3 (weakly tense) in week 1, while in Kessler and Turner's study, 75% of the cats registered this median score or higher. In week 2, none of the cats in the current study had a median score of above 3, while 35% of Kessler and Turner's cats did. This could be a function of the small sample size in the current study, or differences in methodology between the current study and those in Kessler and Turner's, or it could simply be that the cats in the current study were less stressed.

3.5.3 Faecal glucocorticoid metabolites

Interestingly, the FGM data in the current study found that while the physiological stress response declined over time, the only significant difference between weeks was between week 1 and 5. Any subsequent significant between-week differences may have been diminished by the small sample size and high degree of individual variation. There was a steep mean drop in this variable from week 1 to week 2, followed by a relatively stable period for weeks 2-4 and a slight drop in week 5. Hawkins et al. (2004) found a significant decline in cortisol to creatinine ratios across the first 8 days in a shelter, using a sample size of 23 singly housed cats. The statistical analysis in the current study did not take into account the effect of which plate each sample was on. As samples were run sequentially, it is possible that differences between plates created differences between weeks. Unfortunately as the data analysed were weekly averages (and in some cases, daily averages) per cat, it was impossible to include which plate a value came from in the model. However, as the inter-assay coefficient of variation was well within normal laboratory standards, it is unlikely that the effect of which plate each sample came from was substantial.

3.5.4 Post-habituation time budget

This study also provided a post-habituation time budget for singly housed caged cats. Watanabe et al. (2005) have published a time budget for free-ranging cats, and Barry and Crowell-Davis (1999) reported a time budget for cats housed in a home environment, but obviously, the behaviours exhibited by these populations will be different from their caged counterparts.

Ottway and Hawkins (2003), Iki et al. (2011) and Gouveia et al. (2011) have published frequency and/or duration data on the behaviours exhibited by caged cats, but these studies were conducted on communally housed cats. Eckstein and Hart (2000) produced a time budget for singly housed cats, but did not allow enough time for habituation to occur. The time budget data for cage use and posture are novel and provided valuable information. Cats spent the highest percentage of time on the shelf – although this was not significantly different from expected based on surface area. Smith et al. (1994) also reported that communally housed cats used raised structures significantly more than other provided surfaces. The results of these studies suggest that raised platforms are a resource of value to the cat, and that its inclusion in enclosure design may be important.

All the above studies which have published a time budget for cat behaviour have reported that cats spend the largest percentage of their time resting, although at much lower magnitudes than the current study. They also reported much larger percentages of other behaviours, notably those social or active in nature. This suggests that in the current study, cage size and lack of access to conspecifics may have inhibited these behaviours, thereby inflating the percentage of time spent resting.

3.5.5 Issues

Although there was adequate inter-observer agreement for the duration data, inter-observer agreement for the frequency data was inadequate. This was likely due to an oversight when programming the configuration of the data collection. Whenever a cat changed its posture, location, or activity all were registered as changing. This means that if a cat started an observation as ‘standing, on shelf, grooming’ and then changed to ‘sitting, on shelf, grooming’ that would be counted as two instances of grooming, when in reality it was only one. In future projects this should be easy to reconfigure. For the current project, it would be possible to comb through the data files and clean them by creating three separate files for each observation (one for each category: posture, location, and activity) and eliminating all sequential repeats. However, as this would be tremendously time consuming, and there is a large amount of data presented even without fixing and presenting the frequency data, this has not been carried out.

3.5.6 Conclusions

Although the sample size was small and findings must be interpreted with caution, the results indicate that the percentages of times spent per day grooming and eating might have been related to the initial stress (positive and negative relationships respectively) of caging. However, comparison with results of other published studies indicates that the relationship between grooming and stress might not be well defined. The quantitative behavioural results and the FGM results lend support to the CSS as a method of assessing stress in the domestic cat. Furthermore, the results of the quantitative behavioural observations and the CSS indicate that the cats experienced an initial stress response to caging, but that after the first week, habituation to the new environment was more or less complete. Log_e FGM data showed a reduction across time, but the marked reduction from week 1 to week 2 was not statistically significant. This may have been due to a small sample size. As in other environments, cats spent the highest percentage of their time resting. However, compared with other studies of cats in a home environment or caged communally, cats in this study spent more of their time resting, which may indicate that the cage size or lack of access to conspecifics may have inhibited the performance of other behaviours. The results of this and other studies suggest that structures in the cage such as raised platforms may be a valuable resource for the cat, and that its inclusion in enclosure design may be important.

3.6 References

Barry, K.J., Crowell-Davis, S.L., 1999. Gender differences in the social behavior of the neutered indoor-only domestic cat. Appl. Anim. Behav. Sci. 64, 193-211.

Canadian Federation of Humane Societies, 2010. Animal Shelter Statistics. 2012.

Carlstead, K., Brown, J.L., Strawn, W., 1993. Behavioral and Physiological Correlates of Stress in Laboratory Cats. Appl. Anim. Behav. Sci. 38, 143-158.

Cooper, J.O., Heron, T.E., Heward, W.L., 2007. Applied Behavior Science, 2nd Ed. Pearson, Upper Saddle River, N.J.

Daly, J.A., Hogg, E., Sacks, D., Smith, M., Zimring, L., 1983. Sex and Relationship Affect Social Self-Grooming. Journal of Nonverbal Behavior 7, 183-189.

DeVries, T.J., Vankova, M., Veira, D.M., von Keyserlingk, M.A.G., 2007. Short Communication: Usage of Mechanical Brushes by Lactating Dairy Cows. J. Dairy Sci. 90, 2241-2245.

Dybdall, K., Strasser, R., Katz, T., 2007. Behavioral differences between owner surrender and stray domestic cats after entering an animal shelter. Appl. Anim. Behav. Sci. 104, 85-94.

Eckstein, R.A., Hart, B.L., 2000. The organization and control of grooming in cats. Appl. Anim. Behav. Sci. 68, 131-140.

Ferguson, C.J., 2009. An effect size primer: A guide for clinicians and researchers. Professional Psychology: Research and Practice 40, 532-538.

Frank, D.F., Erb, H.N., Houpt, K.A., 1999. Urine spraying in cats: presence of concurrent disease and effects of a pheromone treatment. Appl. Anim. Behav. Sci. 61, 263-272.

Fruteau, C., Voelkl, B., van Damme, E., Noe, R., 2009. Supply and demand determine the market value of food providers in wild vervet monkeys. Proc. Natl. Acad. Sci. U. S. A. 106, 12007-12012.

Gourkow, N., Fraser, D., 2006. The effect of housing and handling practices on the welfare, behaviour and selection of domestic cats (*Felis sylvestris catus*) by adopters in an animal shelter. Anim. Welfare 15, 371-377.

Gouveia, K., Magalhães, A., de Sousa, L., 2011. The behaviour of domestic cats in a shelter: Residence time, density and sex ratio. Appl. Anim. Behav. Sci. 130, 53-59.

Griffith, C.A., Steigerwald, E.S., Buffington, C.A.T., 2000. Effects of a synthetic facial pheromone on behavior of cats. J. Am. Vet. Med. Assoc. 217, 1154-1156.

Hawkins, K.R., Bradshaw, J.W.S., Casey, R.A., 2004. Correlating cortisol with a behavioural measure of stress in rescue shelter cats. Anim. Welfare 13, S242-S243.

Hodges, J.K., Heistermann, M., 2003. Field Endocrinology: Monitoring Hormonal Changes in Free-Ranging Primates, in: Setchell, J.M., Curtis, D.J. (Eds.). Cambridge University Press, Cambridge, pp. 353-370.

Iki, T., Ahrens, F., Pasche, K.H., Bartels, A., Erhard, M.H., 2011. Relationships between scores of the feline temperament profile and behavioural and adrenocortical responses to a mild stressor in cats. Appl. Anim. Behav. Sci. 132, 71-80.

Jansen, R., Wiertz, L., Meyer, E., Noldus, L., 2003. Reliability analysis of observational data: Problems, solutions, and software implementation. Behavior Research Methods 35, 391-399.

Kessler, M.R., Turner, D.C., 1997. Stress and adaptation of cats (*Felis silvestris catus*) housed singly, in pairs and in groups in boarding catteries. Anim. Welfare 6, 243-254.

Landis, J.R., Koch, G.G., 1977. The Measurement of Observer Agreement for Categorical Data. Biometrics 33, 159-174.

Macht, M., 2008. How emotions affect eating: A five-way model. Appetite 50, 1-11.

Maestripieri, D., Schino, G., Aureli, F., Troisi, A., 1992. A modest proposal: displacement activities as an indicator of emotions in primates. Anim. Behav. 44, 967-979.

Martin, P., Bateson, P., 2007. Measuring Behaviour : An Introductory Guide, 3rd ed. Cambridge University Press, Cambridge.

Mendl, M., Harcourt, R., 2000. Individuality in the cat: origins, development and stability, in: Turner, D.C., Bateson, P. (Eds.), 2nd ed. Cambridge University Press, Cambridge, Uk, pp. 47-64.

Moberg, G.P., 2000. Biological Response to Stress: Implications for Animal Welfare, in: Moberg, G.P., Mench, J.A. (Eds.). CABI Pub., Wallingford, UK, New York, NY, pp. 1-22.

Möstl, E., Palme, R., 2008. Measuring Faecal Steroid Metabolites with Enzyme Immunoassays (EIA) on Microtitre Plates using Biotinylabled Steroids as Labels. University of Veterinary Medicine, Vienna, Austria, pp. 1-10.

Ottway, D.S., Hawkins, D.M., 2003. Cat housing in rescue shelters: A welfare comparison between communal and discrete-unit housing. Anim. Welfare 12, 173-189.

Pageat, P., Gaultier, E., 2003. Current research in canine and feline pheromones. Veterinary Clinics of North America-Small Animal Practice 33, 187-211.

Palme, R., Möstl, E., 1997. Measurement of cortisol metabolites in faeces of sheep as a parameter of cortisol concentration in blood. Zeitschrift Fur Saugetierkunde 62 Supp/2, 192-197.

Palme, R., Schatz, S., Möstl, E., 2001. [Influence of a vaccination on faecal cortisol metabolite concentrations in cats and dogs.]. DTW (Deutsche Tieraerztliche Wochenschrift) 108, 23-25.

Rochlitz, I., 1999. Recommendations for the housing of cats in the home, in catteries and animal shelters, in laboratories and in veterinary surgeries. J Feline Med Surg 1, 181-91.

Rochlitz, I., 2000. Feline welfare issues, in: Turner, D.C., Bateson, P. (Eds.), 2nd ed. Cambridge University Press, Cambridge, UK, pp. 207-226.

Smith, D.F., Durman, K.J., Roy, D.B., Bradshaw, J.W., 1994. Behavioural aspects of the welfare of rescued cats. The Journal of the Feline Advisory Bureau 31, 25-28.

Tanaka, A., Wagner, D.C., Kass, P.H., Hurley, K.F., 2012. Associations among weight loss, stress, and upper respiratory tract infection in shelter cats. J. Am. Vet. Med. Assoc. 240, 570-576.

Touma, C., Palme, R., 2005. Measuring fecal glucocorticoid metabolites in mammals and birds: the importance of validation. Ann. N. Y. Acad. Sci. 1046, 54-74.

UK Cat Behaviour Working Group, 1995. An Ethogram for Behavioural Studies of the Domestic Cat (*Felis silvestris catus*). Universities Federation for Animal Welfare, Wheathampstead, Herts.

Watanabe, S., Izawa, M., Kato, A., Ropert-Coudert, Y., Naito, Y., 2005. A new technique for monitoring the detailed behaviour of terrestrial animals: A case study with the domestic cat. Appl. Anim. Behav. Sci. 94, 117-131.

Weiss, E., Miller, K., Heather Mohan-Gibbons, Vela, C., 2012. Why Did You Choose This Pet?: Adopters and Pet Selection Preferences in Five Animal Shelters in the United States. Animals, 144.

Wenstrup, J., Dowidchuk, A., 1999. Pet Overpopulation: Data and Measurement Issues in Shelters. J. Appl. Anim. Welfare Sci. 2, 303.

Whitten, P.L., Brockman, D.K., 1998. Recent Advances in Noninvasive Techniques to Monitor Hormone-Behavior Interactions. Yearbook of Physical Anthropology 41, 1.

4 ASSESSING BEHAVIOURAL STYLE, AND DISCRIMINATING BETWEEN BOLD AND SHY DOMESTIC CATS IN A SHELTER SETTING

4.1 Abstract

Whether a cat is bold or shy has the potential to impact its experience in a shelter environment. A simple test to discriminate between bold and shy cats could help shelter workers tailor management practices to the needs of each group. In the current study, a method was established for discriminating between the two modes of this behavioural style. Nine cats were subjected to a 5-min behavioural tests with two quantitative outcome measures, conducted once a week for up to five weeks, depending on residency in the study. Variables which did not change over time, but showed differences between cats, were considered valid as potential measures of temperament. The same test was then conducted in a different environment using seventeen additional cats every other day for up to three days, depending on residency in the shelter. Observer ratings of whether a cat was bold or shy were also conducted in each environment. Data combined from both studies were then analysed to determine which cut-point would correctly classify the greatest percentage of cats overall, as well as the bold and shy cats separately. Latency to emerge from carrier with a cut-point of 10 s was suggested as the most appropriate test for discriminating between bold and shy cats in a shelter setting. It correctly identified a high percentage of cats overall (81%), and bold (74%), and shy (100%) cats separately. In addition, it was quick and easy to administer. Moreover, it was the method best suited to correctly identify shy individuals which were less represented in this population and arguably could derive greater benefit from identification and extra attention.

4.2 Introduction

Temperament traits – often called behavioural styles in the feline literature – are patterns of behaviours exhibited differentially across individuals that are relatively stable across time and context, and have potential to influence the way an individual interacts with their environment (Réale et al., 2007). However, the importance of each particular behavioural style would be variable in different situations. For example, the behavioural style 'sociability' may have less of an impact on a situation that was solitary, than one that required interaction with other individuals.

Within cats specifically, the most commonly studied behavioural style is friendliness towards humans (Durr and Smith, 1997; Feaver et al., 1986; Iki et al., 2011; Lee et al., 1983; Lowe and Bradshaw, 2001; McCune, 1995; Meier and Turner, 1985; Mertens and Turner, 1989; Reisner et al., 1994; Siegford et al., 2003; Turner et al., 1986), which is logical considering the cat is a companion animal. The second most commonly studied behavioural style is bold/shy (Durr and Smith, 1997, McCune, 1995; Reisner et al., 1994), although the definitions provided for both of these behavioural styles vary. Using McCune's (1995) definition of the bold/shy behavioural style – general response to novelty irrespective of whether the novelty is human or object – we could consider any of the studies that investigated friendliness towards humans as also investigating boldness /shyness. In fact, Meier and Turner (1985) and Mertens and Turner (1989) even called the behavioural style they were investigating shy/trusting, which has obvious parallels between the behavioural styles friendliness towards humans and bold/shy. Other behavioural styles investigated in cats include alert and equable (or, even tempered) with other cats (Feaver et al., 1986). The terminology used outside of the feline discourse is similar in many respects, but there are subtle yet important differences in the traits examined and how they are defined (summarised in Réale et al., 2007).

As shelters are potentially stressful environments for a number of reasons including exposure to unfamiliar people, animals, and environment (Carlstead et al., 1993), arguably, of the four behavioural styles commonly studied in cats, bold/shy would be the key behavioural style to mediate a cat's response to the stressors implicit in being housed in a shelter. De Palma et al. (2005) liken bold and shy behavioural styles to Koolhaas et al.'s (1999) proactive and reactive styles, of which the latter was found to be more cautious towards novel external stimuli and mount a higher stress response to stimuli. As two of the primary stressors of shelter environments are exposure to unfamiliar people and environments – and the definition of the bold/shy behavioural style was a general response to novelty irrespective of if the novelty is human or object – it follows that bold cats and shy cats would have greatly different experiences in shelter settings, and could potentially benefit from different management and husbandry conditions. Additionally, there is evidence from the literature that this behavioural style is discrete, meaning that cats were either bold or shy as opposed to existing somewhere on a continuum (Meier and Turner, 1985; Mertens and Turner, 1989). Therefore, it could be advantageous for shelter staff to be able to measure this behavioural style, and classify cats as either bold or shy. Many authors have attempted to measure and discriminate between bold and shy cats, however these methods were often time consuming – i.e. behavioural tests taking from approximately an hour (Lee et al., 1983) to a full summer (Meier and Turner, 1985) – or required intimate knowledge of the cat (Feaver et al., 1986; Turner et al., 1986). While the results of these studies are often shown to be reliable and valid, shelter staff are busy and cat intake rates can be quite high. If a test is to be used in shelter settings by staff to determine what management and husbandry conditions to employ for each cat, an ideal test for this behavioural style would have to be quick and easy, while maintaining rigorous standards for reliability and validity.

The purpose of this study was to identify a quick and easy test, for use in research studies and in shelters, to determine whether a cat is bold or shy. Two potential behavioural measures were examined for their agreement with observer ratings, reliability over time, and ability to identify the best cut-point to use to discriminate between bold and shy cats. These measures were examined in two different environments, and then their findings were combined for analysis.

4.3 Materials and methods

4.3.1 Subjects

Environment 1

Nine domestic cats were obtained from a local animal shelter, of which four were intact males and five were females of unknown reproductive status, all of which tested negative for feline immunodeficiency virus and feline leukemia virus. The average length of stay (LOS) in the shelter before transport approximately 5 km to the study site was 2.8 ± 1.88 (± SD) days, and average estimated age of cats included in this study was 2.1 ± 1.05 years. Three of the cats had been surrendered by their owner and six of them were strays. Due to inappetence or upper respiratory tract infection, three of the cats did not stay in the study for the duration of the five-week study, conducted over seven weeks. As data collection from these cats was not complete, they are included in the analysis only when indicated. At any one time, there were between 3 and 6 cats in the study.

Environment 2

Seventeen cats were tested at the local animal shelter, of which 9 were intact males and 8 were females of unknown reproductive status. The average estimated age of the cats included in this study was 3.6 ± 1.50 years. All of the cats were strays. Average LOS in this environment was 9.0 ± 8.02 days. The LOS was longer than in environment 1, so that the shelter workers would be more familiar with the cats and would be therefore better suited to provide ratings of their behavioural style. The study was conducted on three separate days.

4.3.2 Housing and management

Environment 1

The cats were housed and managed as described in Ch 3, this text.

Environment 2

Cats were housed at the local animal shelter under their standard conditions. As cage size and furnishing varied – and it is likely that food provided, feeding times, level of human interaction, and timing of light cycles did as well – no attempt was made to characterise these conditions.

4.3.3 Behavioural tests

Environment 1

Cats were subjected to one behavioural test every seven days for five weeks, beginning the morning of their first full day in the facility. The order in which cats were tested was randomised. The test was conducted in a circular open-field arena, which was 1.22 m tall with a circumference of 9.75 m, an area of 7.56 m^2, and a small Kennel Cab II cat carrier (Petmate, Arlington, Texas, USA) (48.3 cm x 31.8 cm x 25.4 cm) opening into the arena through the side.

Cats were place in the carrier one at a time, the door was opened, and each cat's behavior was monitored for 5 min. The latency to emerge from the carrier and the percentage of time spent in the carrier was measured. Emerging from the carrier were said to have taken place when >50% of the cat's body had completed the action, and the cat was free to return inside the carrier at any time during the 5 minutes. All tests were video recorded with a CCTV camera (Panasonic, Germany) suspended from the ceiling of the test room, and analysis of behaviour was conducted using Noldus Observer 5 (Noldus Information Technology, Wageningen, The Netherlands). Complete results were obtained from six cats from all five weekly tests, while results from one cat were obtained from the first two weeks and results from two cats were obtained for only the first week.

Environment 2

Cats were subjected to one behavioural test every two days and this was repeated up to two times. The cat was placed into a carrier which was placed in the middle of an empty rectangular room (3.66 m x 2.90 m). The carrier door was opened and the cat was left alone for 5 min. The cat's behaviour was observed through a partially translucent window (1.60 m x 0.48 m), and latency to emerge (>50% of its body) and percentage time spent in the carrier were recorded. Due to events such as transfers or adoptions, data from all three test days were collected from only 10 cats.

4.3.3 Observer ratings

Environment 1

Upon completion of the data collection for all cats, two investigators who were both quite familiar with the animals rated each cat as either bold or shy, based largely on the cat's reaction to novelty (irrespective of whether the novelty is human or object), as described by McCune (1995). Bold individuals were said to be more likely to react sociably or relaxed, and less likely to react distressed or threatened.

Environment 2

The investigator and four shelter attendants who were all familiar with the cats were asked to rate the cats as either bold or shy, using the same definition as in environment 1. This was conducted after the final behavioural test was carried out, but only the investigator was aware of how each cat scored.

4.3.4 Statistical analyses

Environment 1

Behavioural tests

In order to use a test to discriminate between modes of this behavioural style, measures need to be sensitive enough to show differences between individuals, but not be significantly different across weeks – as cross-time consistency is a key feature of temperament. Differences between individuals and across time were analysed using two-way ANOVAs with a random effect of cat and a fixed effect of week for each outcome measure. The natural log was used to transform the measure' latency to emerge', while the arcsine square root used to transform the measure 'percentage of time spent in carrier. Measures were considered to be sensitive to individual differences if cat was a significant factor, and consistent across time if week in study was not a significant factor. Non-significant time effects were confirmed with repeated measures ANOVA (STATA statistical software, version 12.1). Only cats who participated in all weekly test batteries were included in this analysis. In addition, an intraclass correlation coefficient (ICC) (Lessells and Boag, 1987) was computed for purposes of completeness and as a descriptor, whenever its underlying assumptions could be met. Values of ≥ 0.5 were taken to indicate consistent individual responses (Gibbons et al., 2009).

Observer ratings

Dichotomous ratings were compared between observers for agreement using percentage of agreement, and Cohen's kappa. Minimum acceptability of results was set at ≥80 for percent of agreement (Cooper et al., 2007), and ≥0.60 (meaning substantial or better) for Cohen's kappa test statistic (Landis and Koch, 1977).

Environment 2

Behavioural tests

Analyses were conducted on both measure to ensure that they showed significant between-cat differences and non-significant between-week differences. A 2-way ANOVA was used to investigate a $\log_e$ transformation of latency to emerge, while Friedman's test was used to investigate percentage of time spent in carrier. These analyses were restricted to the 10 cats present on all three testing days.

Observer ratings

Agreement between observer ratings was assessed with percentage of agreement and Fleiss's kappa for multiple raters on all 17 cats. Minimum acceptability of results scores were the same as set for environment 1, with the same standards for Fleiss' kappa as for Cohen's kappa. For subsequent analyses, cats were assigned the designation as either bold or shy, based on their most frequent rating. Since the researcher was privy to the results of the behavioural test at the time of observer rating and the shelter workers were not, the agreement was also calculated among the shelter workers alone. This was compared to the overall agreement to assess if the ratings of the researcher were being biased by the results of the behavioural test.

Analysis of combined data from environments 1 and 2

Since neither environment showed the outcome variables 'latency to emerge from carrier' or 'percentage of time spent in carrier' to change over time, the remaining analyses were conducted on data from the first day of testing only. This maximised the sample size by allowing inclusion of the three cats from environment 1 that did not participate in the full set of weekly tests, and the seven cats from environment 2 that were not present for all three test days. Therefore the remaining analyses conducted on latency to emerge and percentage time spent in carrier used data from 26 cats, nine from environment 1, and 17 from environment 2, and the observer ratings associated with each.

Determining the best cut-point

In order to determine whether a cat was bold or shy from the continuous variable latency to emerge, a cut-point had to be determined to discriminate between the terms. Using the observer ratings as the reference test, calculations were conducted to determine the percentage of correct classification at each of the potential cut-points observed, both overall and for bold and shy cats separately (similar to sensitivity and specificity calculations in diagnostic testing). The same analysis was also conducted for the variable percentage of time spent in carrier.

General settings for statistical analysis

All analyses of behaviour in both studies were conducted using Minitab® 15 statistical software (State College, Pennsylvania, USA), unless otherwise stated. Significance level was set at $P<0.05$.

4.3.3 Ethical approval

This project was approved by UPEI's Animal Care Committee under protocol number 09-051, and followed the guidelines of the Canadian Council of Animal Care's "Guide to the Care and Use of Experimental Animals".

4.4 Results

4.4.1 Environment 1

Behavioural tests

Log_e latency to emerge was found to vary significantly across cats ($F_{5,20}=11.70$, $P<0.001$), but not test days ($F_{4,20}=1.89$, $P=0.152$). The arcsine square root of percentage of time spent in carrier was found to vary significantly across cats ($F_{5,20}=13.43$, $P<0.001$), but not test days ($F_{4,20}=1.13$, $P=0.371$). ICCs showed consistent individual responses for both measures (ICC=7).

Observer ratings

The bold/shy behavioural style was found to be reliably assessed between observers, scoring acceptably in the percentage of agreement (89%, SE=10.48) and Cohen's kappa (0.78, SE=0.20, P=0.01). Five cats were identified as bold, while the remaining four were identified as shy for subsequent analyses.

4.4.2 Environment 2

Behavioural tests

Log_e latency to emerge was found to vary significantly across cats ($F_{9,18}$=6.13, P=0.001), but not test days ($F_{2,12}$=0.83, P=0.453). Percentage of time spent in carrier was found to vary significantly across cats (S=19.75, df=9, P=0.020), but not test days (S=0.69, df=2, P=0.710).

Observer ratings

Percent of agreement between observers was 65%, with 11 of 17 cats registering perfect agreement between all five observers. While this number is lower than the 'acceptable' standard set out in the methods section, a lower number is arguably acceptable in environment 2, because there were a higher number of observers than in environment 1, and this value only accounts for cats where there was perfect agreement between all observers. The remaining six cats had agreement between four of the five observers (the dissenting observer varied each time). Fleiss' kappa of agreement between the researcher and the shelter workers was 0.62, SE=0.08, and the Fleiss' kappa between the shelter workers alone was 0.55, SE=0.10. Fourteen cats were identified as bold while the remaining 3 were identified as shy.

4.4.3 Analysis of combined data from environments 1 and 2

Determining the best cut-point

For the variable latency to emerge, the cut-point which correctly classified the greatest percentage of cats overall was 33 s. Two other cut-points were identified for further investigation due to the high percentage of cats correctly classified, they were: 10 s and 30 s. Ten s was chosen as it represented a second peak in percentage of cats correctly classified, and 30 s was chosen as it is an intuitive, round number for testers to use as a cut-point. For the variable percentage of time spent in carrier, the cut-point which correctly classified the greatest percentage of cats was 12%. No other cut-points were investigated for this variable, as there was only one peak in the percentage of cats correctly classified and the closest intuitive round number (10%) lowered the percentage of cats correctly classified to 80%. Both the percentage of time spent in carriers and latencies to emerge from carriers were relatively unimodally distributed. The percentages of bold and shy cats classified correctly by the various cut-points assessed for each outcome measure can be seen in Table 4.1.

Table 4.1 Percentages of correct classification for both bold and shy cats using the selected cut-points

Outcome measure	Cut-point	Total percentage classified correctly	Percentage bold classified correctly	Percentage shy classified correctly
Latency to emerge	33 s	92	100	71
Latency to emerge	10 s	81	74	100
Latency to emerge	30 s	88	95	71
Percentage time spent in carrier	12	85	90	71

4.5 Discussion

The primary goal of this study was to identify an effective and practical test to determine whether a cat was bold or shy, in order to help shelter staff to make informed husbandry or enrichment decisions especially when resources or space are limited. Using the observer ratings as a reference test, whether or not a cat had emerged from its carrier into an empty arena/room with a 33 s cut-point correctly classified the greatest percentage of cats. However, there were far fewer shy cats represented in the sample population than there were bold cats, and these cats could arguably benefit from environmental enrichment and tailored husbandry programs more than bold cats. For these reasons it may be more important to correctly identify shy cats, while still correctly classifying a relatively high percentage of bold cats. As latency to emerge with a 10 s cut-point correctly identified 100% of shy cats, while correctly classifying

74% of bold cats, and >80% of cats overall, this test and cut-point is recommended as the most appropriate to discriminate between bold and shy cats in this scenario. This 10 s test is also very practical time-wise. However, the need for an empty and quiet room for testing is essential, and may in practice be a limitation.

In this study, 2-way ANOVAs – including both cat and week as factors – were used to identify VOI that were different between individuals, but not across time. When assumptions of the model could not be met, Friedman's tests were employed. Often this comparison is conducted using the ICC, which takes into account variation over time of each individual. However, as it works within an ANOVA framework, there are several assumptions implicit in using this statistic. As many of the outcome variables being analysed here could not be transformed to comply with these assumptions, using this test with this data set was impractical. The 2-way ANOVA/Friedman's test method was superior to the ICC for investigating change over time for this particular data set for two additional reasons. First, as there are only 6 cats included in this analysis, estimating variance may not be appropriate to represent a population. Second, the ICC only measures changes in variance over time, and not changes in absolute values. As it is ideal to identify variables for which the first measure of a test is not significantly different than a measure of the same test taken weeks later – to identify usable cut-points for use in shelter settings – the 2-way ANOVA/Friedman p-value seemed more appropriate.

One of the premises of this research relies on the behavioural style bold/shy being expressed as one of two extremes – as was suggested by previous research (Meier and Turner, 1985; Mertens and Turner, 1989). However, the distribution of values in the behavioural test chosen to be most effective at discriminating between the dichotomous categories was not decidedly bimodal. A relatively unimodal distribution of scores on both the latency to emerge and percentage of time spent in carrier tests were observed, suggesting that perhaps this behavioural style does exist on a continuum after all, as has been described in other species (Wilson et al., 1994). It will be interesting to see whether – if future tests do show certain types of environmental enrichment or husbandry techniques benefit either bold or shy cats more profoundly – the effect of these interventions is expressed uniformly by cats on each side of the behavioural style, or if the magnitude of these effects varies along the same continuum as the scores on the behavioural tests.

Finally, discussion of the first objective of this study raised the issue that dichotomising a variable may not always be appropriate, and that even if individuals do classify as one of two extremes of a behavioural style, it is possible that it is more important to correctly classify one type than another. These issues highlight the importance of understanding the specific of the research question, and the intricacies of each data set. Although this study set about to develop a methodological framework that future researchers could use to evaluate (and potentially dichotomise) behavioural styles in other species and/or circumstances, it may not be appropriate for all research questions – although adaptation may be possible in many instances.

4.5.1 Conclusion

Latency to emerge from a carrier and percentage of time spent in a carrier failed to show significant differences across time, but did show significant differences between individuals - therefore showing potential to represent a behavioural style. Observer ratings of the behavioural style bold/shy were shown to have high inter-rater agreement. When comparing the results of these behavioural tests with these observer ratings it was determined that the most appropriate test for quickly dichotomising between bold and shy cats in a shelter setting was placing a cat in a carrier inside a quiet empty room, and observing whether or not > 50% of its body emerges in less than 10 s. Any cat that emerges during this period can be classified as bold, and any cat that fails to do so can be classified as shy. There is potential for the framework used for developing this test in the current study to be adapted to dichotomise between modes of behavioural styles for other species in other environments.

4.6 References

Carlstead, K., Brown, J.L., Strawn, W., 1993. Behavioral and Physiological Correlates of Stress in Laboratory Cats. Appl. Anim. Behav. Sci. 38, 143-158.

Cooper, J.O., Heron, T.E., Heward, W.L., 2007. Applied Behavior Science, Second Edition ed. Pearson, Upper Saddle River, N.J.

De Palma, C., Viggiano, E., Barillari, E., Palme, R., Dufour, A.B., Fantini, C., Natoli, E., 2005. Evaluating the temperament in shelter dogs. Behaviour 142, 1307-1328.

Durr, R., Smith, C., 1997. Individual differences and their relation to social structure in domestic cats. Journal of Comparative Psychology 111, 412-418.

Feaver, J., Mendl, M., Bateson, P., 1986. A Method for Rating the Individual Distinctiveness of Domestic Cats. Anim. Behav. 34, 1016-1025.

Gibbons, J., Lawrence, A., Haskell, M., 2009. Responsiveness of dairy cows to human approach and novel stimuli. Appl. Anim. Behav. Sci. 116, 163-173.

Iki, T., Ahrens, F., Pasche, K.H., Bartels, A., Erhard, M.H., 2011. Relationships between scores of the feline temperament profile and behavioural and adrenocortical responses to a mild stressor in cats. Appl. Anim. Behav. Sci. 132, 71-80.

Koolhaas, J.M., Korte, S.M., De Boer, S.F., Van Der Vegt, B.J., Van Reenen, C.G., Hopster, H., De Jong, I.C., Ruis, M.A.W., Blokhuis, H.J., 1999. Coping styles in animals: current status in behavior and stress-physiology. Neurosci. Biobehav. Rev. 23, 925-935.

Landis, J.R., Koch, G.G., 1977. The Measurement of Observer Agreement for Categorical Data. Biometrics 33, pp. 159-174.

Lee, R.L., Zeglen, M.E., Ryan, T., Hines, L.M., 1983. Guidelines: Animals in nursing homes. California Veterinarian 3, 22a-26a.

Lessells, C.M., Boag, P.T., 1987. Unrepeatable repeatabilities: A common mistake. Auk 104, 116-121.

Lowe, S.E., Bradshaw, J.W.S., 2001. Ontogeny of individuality in the domestic cat in the home environment. Anim. Behav. 61, 231-237.

McCune, S., 1995. The Impact of Paternity and Early Socialization on the Development of Cats Behavior to People and Novel Objects. Appl. Anim. Behav. Sci. 45, 109-124.

Meier, M., Turner, D.C., 1985. Reactions of house cats during encounters with a strange person: Evidence for two personality types. Journal of the Delta Society 2, 45-53.

Mertens, C., Turner, D.C., 1989. Experimental analysis of human-cat interactions during first encounters. Anthrozoös 2, 83-97.

Réale, D., Reader, S.M., Sol, D., McDougall, P.T., Dingemanse, N.J., 2007. Integrating animal temperament within ecology and evolution. Biol. Rev. 82, 291-318.

Reisner, I.R., Houpt, K.A., Erb, H.N., Quimby, F.W., 1994. Friendliness to humans and defensive aggression in cats: the influence of handling and paternity. Physiol. Behav. 55, 1119-1124.

Siegford, J.M., Walshaw, S.O., Brunner, P., Zanella, A.J., 2003. Validation of a temperament test for domestic cats. Anthrozoos 16, 332-351.

Turner, D.C., Feaver, J., Mendl, M., Bateson, P., 1986. Variation in domestic cat behaviour towards humans: A paternal effect. Anim. Behav. 34, 1890-1892.

Wilson, D.S., Clark, A.B., Coleman, K., Dearstyne, T., 1994. Shyness and Boldness in Humans and Other Animals. Trends in Ecology & Evolution 9, 442-446.

5 ENVIRONMENTAL ENRICHMENT CHOICES OF DOMESTIC CATS

5.1 Abstract

The choices of cats for different types of environmental enrichment (EE) were investigated in a choice test. It was hypothesised that cats would spend more of their time in an enriched compartment than an empty compartment, and that they would consistently spend more of their time in proximity to a specific type of EE. Twenty-six cats were kept singularly in choice chambers for 10 days. Each chamber had a central area containing food, water, and a litter tray, and four centrally-linked compartments containing different types of EE: 1) an empty control, 2) a play opportunity with a prey-simulating toy, 3) a shelf to provide a perching opportunity, and 4) a cardboard box with a hole in the side to provide a hiding opportunity. Cat movement between compartments via 'cat-flap' doors was monitored using a data-logger. Results showed that each enriched compartment was visited significantly more during the light period, indicating that the cats were more active during the day. Median number of visits per day differed significantly between compartments: 4 (IQR=4) empty compartment, 3 (IQR=4) toy compartment, 3 (IQR=5) perching compartment, and 5 (IQR=5) hiding compartment. Median percentage of time spent in each compartment differed significantly between compartments: 3% (IQR=6) empty compartment, 2% (IQR=9) toy compartment, 6% (IQR=32) perching compartment, and 55% (IQR=46) hiding compartment. Pairwise comparisons showed that cats spent significantly longer per visit and spent a higher percentage of time in the hiding compartment than in the empty compartment and the toy compartment, but the former was not significantly different from the shelf compartment. The results suggest that a hiding box may be a resource of value to cats, and that its inclusion in enclosure design could be important. Further investigation of the behavioural and physiological differences between cats provided and not provided with a hiding box is required to substantiate the value of this resource.

5.2 Introduction

Shelters are potentially stressful environments for cats (Rochlitz, 1999; Shepherdson et al., 1993). Within a shelter, cats might suffer from an insufficiently enriched environment in terms of an appropriate social environment (Ellis, 2009) or lack of physical complexity (Rochlitz, 2000; 2005). Some studies have provided evidence that hiding may represent an important strategy for coping with the stressors or fear associated with exposure to the unfamiliar (Carlstead et al., 1993a; Kry and Casey, 2007; Smith et al., 1994). Elevated areas, such as shelves or other platforms have also been suggested as "essential to their well-being" (Rochlitz, 1999) and cats have been observed to use them more, both scientifically (Podberscek et al., 1991; Smith et al., 1994) and anecdotally (Deluca and Kranda, 1992) . These elevated areas have been suggested as vantage points from which cats can monitor their surroundings (Ellis, 2009; Rochlitz, 1999; Smith et al., 1994) and could play a key role in exhibiting vigilance to cope with these novel circumstances. Furthermore, Ellis (Ch 3, this text) identified the shelf as the location within the cage where cats spent the majority of their time, after cats had habituated to a novel environment, lending further credence to the importance of this resource to the cat. Another potential coping mechanism of cats in shelters is the ability to play, or perform prey-catching behaviours. Wemelsfelder (1997) argued that the lack of performing an inhibited, but highly motivated behaviour may result in a void of activity that is filled with boredom. Therefore, introducing toys – particularly those simulating predation – would help cater to that potential behavioural need and thus reduce the deleterious effects of boredom that can be associated with lack of physical complexity in otherwise barren shelter cages. While play is not generally exhibited when an animal is undergoing prolonged negative emotions (Spinka et al., 2001), such as those that may be experienced in captivity, it has been predicted that after an initial appraisal period and subsequent assessment that there is no immediate danger, the frequency of play should increase in novel habitats (Spinka et al., 2001). In the literature on play in humans, it was found that distressed preschoolers on the first day of school in a play treatment group became less anxious than those in a treatment group that were read a story (Barnett, 1984). The introduction of toys to singly housed cats was found to reduce inactivity by de Monte and Le Pape (1997), and the authors interpreted this reduction of inactivity as an indicator of improved psychological well-being. Furthermore, in an attempt to determine toy preference in cats, Denenberg (2003) compared the reactions of cats to 10 different toys of varying characteristics. Results indicated that the cats chose toys stimulating chase/predation, and containing food.

It is also possible that individuality in the cat may affect their interaction with confinement, stress, and enrichment. McCune (1992) reported two different styles of behavioural responses to stress – passive and active – each of which are reported to respond differently to confinement (summarised in Ellis, 2009). It has been suggested that different types of environmental enrichment may be more appropriate for these different types of responders (Ellis, 2009). For example, shy cats may rely on hiding opportunities in order to cope with the stress of exposure to unfamiliar people, animals, and environments, while bold cats may actively engage with prey simulating toys in order to stave off the boredom that can be associated with lack of physical complexity in otherwise barren shelter cages.

If in a shelter setting, appropriate enrichment matched to the behavioural style of the cat can be provided, it is likely that the welfare of the cats will be improved. One of the main approaches to study animal welfare is a feelings-based approach, which considers the subjective experiences of an animal, making an attempt to reduce negative experiences and promote positive ones (Fraser et al., 1997). Since the feelings of these animals cannot be observed directly, researchers have developed certain techniques to assess them, such as conducting choice tests. In this test, individuals are presented with two or more options, and they are required to choose one of them. Behavioural choices can be 'measured' by assessing how an animal allocates its time in a resource rich environment, using units such as latency to approach, or frequency of access. However, as summarised by Duncan (1992) and Fraser and Nicol (2011), it cannot be assumed that the choices made by an animal are a straight forward indication of what is best for its welfare. Instead this test has been recommended as more of a starting point into the investigation of the feelings of an animal (Duncan, 2005), and that subsequent investigation into either the motivation to access the resources which animals access more often or spend more time with – often called a consumer-demand test (Dawkins, 1983) – or the effects of providing or denying the resources which animals access more often or spend more time with should then be performed (Fraser and Nicol, 2011) to evaluate the relationship between these resources and the welfare of the subjects.

The goal of this study was to investigate how shelter cats allocate their time between biologically relevant types of environmental enrichment in a choice test. The relationship between behavioural style and visit frequency/time allocation, and differences in visit frequency/time allocation exhibited in the light and dark period were also examined.

5.3 Methods

5.3.1 Study subjects

This study was conducted using 26 domestic cats obtained from a local animal shelter. All procedures were conducted on batches of three cats at a time, except in one instance in which only two suitable cats were obtainable. Fourteen cats were male males (13 were intact and 1 was neutered) , 12 were females (of unknown reproductive status), mean length of stay in the shelter before transport approximately 5 km to the study site was 6 ± 8.3 days (±SD), and mean estimated age of the cats included in this study was 2.5 ± 1.2 years. Seven of the cats had been surrendered by their owner and 19 of them were strays.

5.3.2 Housing and care

Directly after arrival at the study site, cats were subjected to a standard veterinary exam, and then placed singly in plus-shaped enclosures in a dedicated research room. Rooms were kept at 21^{o}C, and illuminated using a 12:12 light cycle, with going lights on at 06:00 h and off at 18:00 h each day. Cats were fed 100g Nutrience Holistic indoor cat food (Hagen, Baie d'Urfé, QC, Canada) once daily: between 11:00 h and 12:00 h. Wet food or treats were also used for training, as discussed below. Each day all remaining food was removed, and new food was provided. Remaining food was weighed in order to monitor food intake. Whenever intake was low for multiple days, diets were supplemented with dry Purina Friskies or cans of wet Royal Canin Medi Cal Recovery or Hills Prescription Diet Critical Care. Additionally, cats were weighed on entrance, after 5 days, and at the end of the study.

5.3.2 Choice chamber description

The enclosure design was modeled after Blom et al. (1992) and Seaman et al. (2008). Cats were housed singly in one of three plus-shaped enclosures, with five separate compartments. The entire enclosure will henceforth be called a 'choice chamber'. The centre compartment contained food and water dishes and a 13 cm x 36 cm x 31 cm litter box filled with approx 450 g of Yesterday's News cat litter (Purina, St. Louis, Missouri, USA). Each external compartment contained a different type of environmental enrichment. Compartment 1 (CTRL) contained no additional enrichment items and served as a control compartment. Compartment 2 (TOY) contained a Mouse Chaser (Ethical Pet, part number 2730, Bloomfield, New Jersey, USA) which is a small round prey-simulating toy, which houses an artificial mouse that runs around a track when the cat activates a motion sensor. These toys were modified so that they were powered by a 12 V power supply, as preliminary testing showed that batteries would run out after five hours of continual use, and during a 10-day study period it was possible that multiple battery replacements could be necessary. Compartment 3 (SHELF) contained one 44.5 cm x 61 cm x 30 cm shelf of a resin shelving unit (Kis, Cod. 009503, Milton, Ontario, Canada) for perching. Finally, compartment 4 (BOX) contained a 76.2 cm x 35.5 cm x 35.5 cm cardboard hiding box, with a half-oval hole cut in the front measuring 26 cm in width, and 18 cm in height. Cats were able to go inside this box, but not to perch on top of it. The frame of the choice chamber was made of white, 1/2 inch PVC tubing connected with three-ended or five-ended PVC elbows. Around the exterior of the frame, green vinyl fencing (Peak Fencing, model number 3406, Richmond, British Columbia, Canada) was secured using cable ties. The walls separating the outside compartments from the central compartment were made of 3 mm thick, clear poly acrylic sheets cut to 76.5 cm x 81.5 cm, which were attached to the PVC tubing with cable ties. The top of each compartment was fitted with a lid made of PVC tubing attached with four, two-ended PVC elbows. Green vinyl fencing was stretched across the frame. Lids opened away from the centre and were attached on that side with cable ties. The centre lid simply rested on top. When lids were closed they were fastened with pipe cleaners. Each compartment of the choice chamber was 80 cm x 80 cm x 80 cm and had a 76 cm x 76 cm polyester PetStore pet blanket attached with cable ties on the bottom. All poly acrylic sheets connecting outside compartments to the centre compartment had two cat flap doors (PetSafe, model P1-4W-11, Indianapolis, Illinois, USA) installed 15 cm from the bottom of the glass, 15.5 cm from either side, and 16 cm from each other. All cat doors were fitted with a magnet on the flap, and a normally open reed switch (Hamlin, DRR-129, Montreal, Quebec, Canada) on the frame. Wires were connected to the reed switches, glued and taped to the poly acrylic sheets, and fed through the PVC piping terminating at data loggers outside of the enclosure. The reed switches from both cat doors leading to the same external compartment were connected to a single data logger – the HOBO U11 3 State/1 Event Data Loggers - U11-001 (Onset, Bourne, Massachusetts, USA). As these loggers were designed to log state changes on three lines, and event changes on the final line, in this study they registered a data point every time a cat door opened or closed in the state lines, and every time a door opened in the event line. In order to make the results from both state and event lines uniform, only door openings were considered in the state lines (data cleaning discussed further in Ch 6, this text). A circular barrier made of 122 cm x 846 cm x 0.4 cm Plaskolite, white corrugated plastic sheeting (Plaskolite, Model 1TW4896A, Columbus, Ohio, USA) enclosed each choice chamber, restricting visual distractions outside of the experimental set up. A CCTV camera (Panasonic, Hamburg, Germany) was suspended from the ceiling above each choice chamber and video was recorded continually using a time-lapse video recorder. Tapes were changed daily at approximately 12:00 h. Two of these enclosures were erected in one room and one in another. Figure 5.1. is a simple schematic of the design. Between batches of cats the choice chambers were raised up off the ground and the floor was cleaned with 1/32 Quatricide PB 13. All of the blankets were laundered, and the food/water bowls and litter pans were put through a cage washer at a maximum cycle temperature of 85^{o}C. Due to the electronic features of the choice chamber, it was cleaned and disinfected by spraying with 1/64 Accel accelerated hydrogen peroxide. The toy and shelf were cleaned in the same way, and the cardboard hiding box was replaced for each cat.

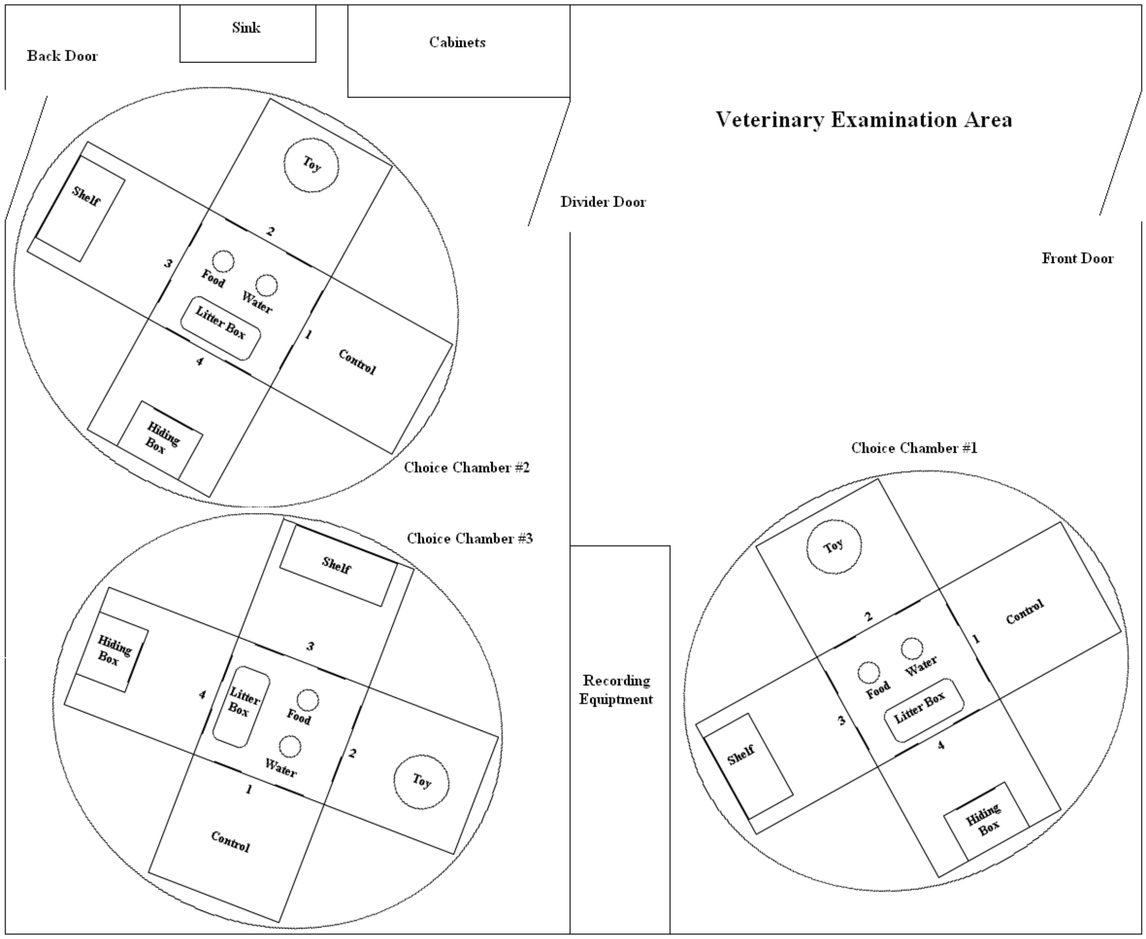

Figure 5.1 Birds-eye view of choice chamber and compartment arrangement (not to scale)

5.3.4 Experimental procedure

Temperament test

After being selected as subjects, cats were subjected to an emergence test to classify them as bold or shy. This test was adapted from what was described in Ch 4, this text. Briefly, cats were placed into a carrier which was placed in the middle of a room. The carrier door was opened and the cat was left alone while its latency to emerge (>50% of its body) was monitored. There were three differences between this version of the test and that used in Ch 4. As the room used in Ch 4 was unavailable, a furnished room rather than an empty room was used. The cats were monitored by keeping the door open a crack (while observation took place through a window in Ch 4). The final difference was that as the cut-point for determining whether a cat was bold or shy was established at 10 s, the test was only conducted for 30 s as opposed to 5 min. Additionally, immediately preceding return of the cats to the shelter, the researcher conducted qualitative observer ratings assigning the cats as either bold or shy based on her subjective experience with the individuals. These ratings were used to validate the findings of the emergence test. Initially the experimenter's ratings were to be checked for agreement against the classification of the Humane Society's veterinarian during the examination. Unfortunately this part of the examination form was rarely filled out, and thus the behavioural style classification based on opinion was based on the experimenter's assessment alone.

Choice chamber test

Following veterinary examination, cats were placed in the centre compartment of the choice chamber at approximately 11:00 h, and the daily food ration was provided. On day one of the experiment, the flaps were absent from the cat doors, allowing cats free access to all compartments of the choice chamber, facilitating familiarity with compartment contents. At approximately 15:30 h the experimenter habituated the cats to her presence by entering the choice chamber, sitting in the centre compartment and attempting to engage the cat for 20 min. If the cat approached, petting and playing took place; if the cat did not approach, the experimenter simply spoke gently to the cat and continued to attempt to engage it. This habituation was conducted in order to facilitate future training. The following day, beginning at around 10:00 h, the flaps were installed in the cat doors and the experimenter trained the cats to use the doors for a 30 min period. Training consisted of attempting to coax the cats through the doors with food rewards or social interaction – some cats responded better to one or the other method. At the end of this period the HOBO data loggers were launched and began collecting data on door use. At approximately 15:30 h deliberate door use was assessed through visual observation or reviewing the activity of the data loggers, and if observed, this was taken as an indication that the cat was adequately trained to use the doors. If the cat was judged not to be able to use the doors deliberately, another 30 min training session was carried out. Training sessions were carried out every morning and afternoon until the cat was judged to be able to use the doors deliberately, at which point the following day was considered their first fully trained day. After this, cats were monitored sporadically to ensure continued door use was performed. At the end of the 10 day period, the cats were removed from the choice chambers, given a veterinary examination, and returned to the local Humane Society. Data were then uploaded from the data loggers. The agreement and reliability of this automated method (HOBO data logger) to record the movement of a cat in and out of a compartment was compared with observation of video recordings. The results of this comparison are described in Ch 6, this text.

5.3.5 Statistics

Temperament assessments using the emergence test and observer rating were compared for agreement using Cohen's kappa. Minimum acceptability was set at >0.6 (or substantial to almost perfect) for Cohen's kappa statistic (Landis and Koch, 1977).

In order to investigate whether cats use the compartments differently, frequency of compartment visit and percentage of time spent in each external compartment (i.e., excluding the centre compartment) was analysed from the data collected by the data-loggers. These parameters were initially computed daily (as total number of visits, and overall percentage of time spent) for each cat and compartment.

First, the effect of time (days since trained) was assessed separately for both of the parameters and each compartment. A Friedman's test with 'Cat ID' blocked by 'Days since trained' was used to analyse the effect of time (days since trained). However, as Friedman's test requires a balanced design, a slight alteration to the data set was necessary. Cat #2 was only successfully trained for three days, and all other cats were trained for a minimum of seven days. Therefore, for this analysis cat #2 was not considered, nor were the data from days greater than seven from any of the other cats.

Next, median values for both parameters were then computed across all days for every cat individually, including cat #2. For both parameters, a Friedman's test blocked by 'Cat ID' was used to compare the compartments, and pairwise comparisons between specific compartments utilised Bonferroni-adjusted sign tests. Effectively, Friedman's test
is based on the within-cat rankings of the four compartments.

Daily compartment use was then separated into light and dark periods, and median values for each cat, compartment, and parameter were calculated. Significant differences in compartment use between light and dark periods were established using sign tests.

Median values for each cat, compartment, and parameter were then used to investigate the effects of sex, whether the cats were strays or surrendered, in which choice chamber they were housed, and whether they were ≤2 or >2 years old, using multifactorial, additive linear models. If needed, the response variables (median parameters) were transformed to meet model assumptions, as evaluated by the residuals. When significant differences between choice chambers were found, Bonferroni-adjusted pairwise comparisons were used to compare between individual compartments.

Median values for both parameters for each cat were then used to look for differences in compartment use (for each compartment separately) between bold and shy cats, using a Mann-Whitney U test. When looking at this factor on its own, linear models were not employed because data could not be transformed to satisfy the implicit assumptions. These analyses were conducted using the dichotomous classification of both methods of temperament assessment (observer rating and the emergence test with a 10 s cut-point).

5.3.6 Ethical approval

This project was approved by UPEI's Animal Care Committee under protocol number 09-051, and followed the guidelines of the Canadian Council of Animal Care's "Guide to the Care and Use of Experimental Animals".

5.4 Results

5.4.1 Temperament test

The emergence test defined 12 cats as shy and 14 cats as bold, while the opinion test defined 11 cats as shy and 15 cats is bold. However, while the tests identified similar numbers of cats as bold or shy, the classification of individual cats varied, resulting in a Cohen's kappa of 0.30 ($P=0.05$) which is only considered fair agreement (Landis and Koch, 1977).

5.4.2 Choice chamber

Effect of time

Cats were considered trained for a median of 9 days (IQR=1), underwent training for a median of 1 day (IQR=1), and spent a median of 6 days (IQR=3) in either the animal shelter or the test facility before they were considered trained.

Time – measured as days since considered reliably trained – was not found to be a significant factor in any of the compartments, using either of the parameters ($P>0.1$). This means that there was no significant difference between days in either the percentage of time spent in, or frequency of visit to, any of the compartments (CTRL, SHELF, BOX, or TOY). Cats used each consistently across time. The daily median, IQR, and N values separated by compartment used for these analyses can be found in Appendix A: Table A.1 for frequency of compartment visit, and Table A.2 for percentage of time spent in each compartment.

Overall visit frequency/time allocation

A Friedman's test revealed a significant difference in daily frequency of visit to each compartment ($S=10.97$, $df=3$, $P=0.01$), however, subsequent Bonferroni-adjusted sign tests revealed no significant differences between specific compartments (Table 5.1). A Friedman's test revealed a significant difference in percentage of time spent in each compartment ($S=23.78$, $df=3$, $P<0.01$), and subsequent Bonferroni-adjusted sign tests revealed that the percentage of time spent in BOX was significantly greater than in CTRL and TOY (Table 5.1). The median percentage of time spent on the SHELF was apparently higher than both CTRL and TOY, suggesting that SHELF may be the next most used compartment. However due to the large degree of individual variation and the correction for multiple testing, these differences were not found to be significant (Table 5.1).

Effect of time of day

Sign tests revealed that frequency of compartment visit was significantly greater in the light period than the dark period for all of the enriched compartments (TOY, SHELF, and BOX) (Table 5.2). Figure 5.2 reveals the largest number of compartment changes occurred between 08:00 h and noon. There were no significant differences between the light and dark periods in the percentage of time spent in any of the compartments (Table 5.2).

Table 5.1 Significant differences between compartments in the median (IQR) daily frequency of compartment visit, and median overall percentage of time spent in each compartment per day (N=26)

Compartment	Frequency of visit	Percentage of time (%)
CTRL	3^{a} (4)	2^{a} (4)
TOY	2^{a} (4)	1^{a} (5)
SHELF	3^{a} (4)	5^{ab} (33)
BOX	5^{a} (4)	52^{b} (50)

^{ab} Within a column, different superscripts indicate a significant difference (P<0.05) using a sign test after Bonferroni-adjustment

For frequency of visit (the only parameter that showed significant differences in compartment use between light and dark periods) a bar chart (Figure 5.2) was created to give a visual representation of how differences were distributed throughout the day. For each hour of the 24 h period, the total number of times a compartment was accessed during the course of the study is represented, and is stacked by compartment.

Table 5.2 Median (IQR) of frequency of compartment visit, and percentage of time spent, per 12 h light and dark period, and significant differences using a sign test (N=26)

	Frequency of visit			Percentage of time (%)		
Comp	Light	Dark	P	Light	Dark	P
CTRL	3 (3)	1 (2)	0.08	3 (8)	1 (6)	0.56
TOY	2 (3)	1 (2)	0.03	2 (10)	10 (13)	0.85
SHELF	3 (1)	1 (2)	<0.01	14 (42)	6 (35)	0.85
BOX	3 (3)	2 (3)	<0.01	63 (39)	69 (66)	0.85

Comp=compartment
P=P-value of sign test between light and dark periods within each measure for each compartment separately

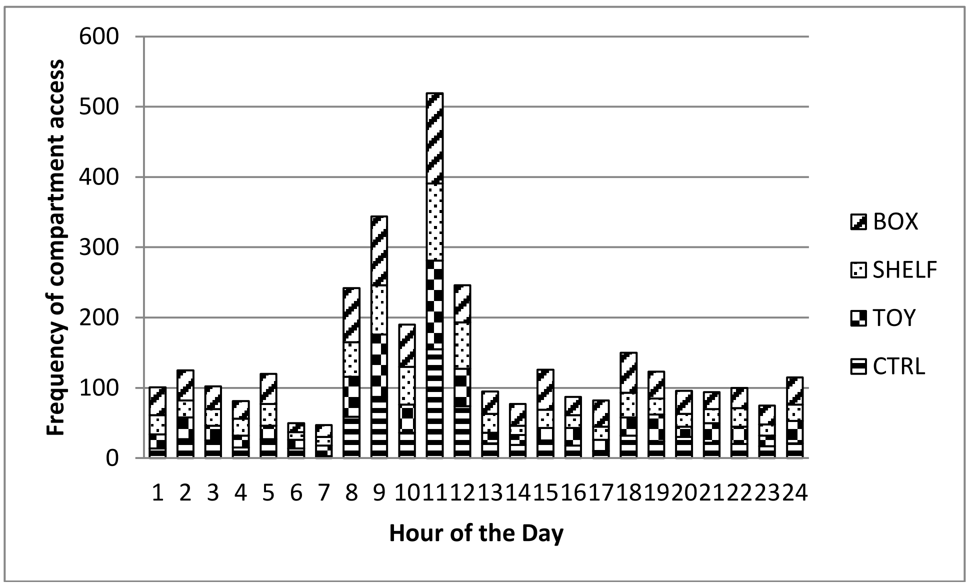

Figure 5.2 Total number of compartment accesses by all cats over the entire study by hour of day for compartments CTRL, TOY, SHELF, and BOX

Effects of cat-level characteristics

There were 14 male and 12 female cats; 19 stray and seven surrendered cats, eight cats were housed in choice chamber 1, eight cats were housed in choice chamber 2, and 10 cats were housed in choice chamber 3; and there were 16 cats ≤2 and 10 cats > 2 years of age. None of the cat-level characteristics investigated here were found to be significant predictors of frequency of visit (Table 5.3) for any of the compartments. Percentage of time spent in TOY or BOX had no significant relationship with any of the cat-level characteristics. Whether cats were stray or surrendered was found to be a significant predictor of the time spent in CTRL; surrendered cats spent significantly more of their time in CTRL than did stray cats. Which of the three choice chambers a cat was housed in was found to be a significant predictor of the percentage of time spent in SHELF; cats in choice chamber 3 spent significantly more of their time in SHELF than did cats in choice chamber 2, whereas choice chamber 1 did not differ significantly from chambers 2 and 3 (Table 5.3).

Table 5.3 P-values for the effects of cat-level characteristics on frequency of visit to, and percentage of time spent in each compartment, obtained by linear models

	Compartment	Sex, df(1,20)		Stray/ Surrender, df(1,20)		Choice Chamber, df(3,20)		Age ≤2 years, df(1,20)	
		F	P	F	P	F	P	F	P
Frequency of visit	CTRL	0.69	0.42	0.54	0.47	2.81	0.08	0.04	0.84
	TOY	0.32	0.58	0.58	0.46	0.56	0.58	0.00	0.99
	SHELF[a]	0.73	0.40	0.15	0.70	2.84	0.80	0.10	0.75
	BOX	0.48	0.50	0.79	0.39	1.06	0.37	0.00	0.95
Percentage of time	CTRL[b]	0.04	0.85	6.67	0.02	0.98	0.39	2.10	0.16
	TOY[b]	0.03	0.87	1.42	0.25	0.26	0.77	0.70	0.41
	SHELF[b]	0.09	0.76	2.73	0.11	4.55	0.02	0.56	0.46
	BOX[b]	0.08	0.78	0.27	0.61	2.14	0.14	0.06	0.81

Specific bold/shy differences

Tests conducted comparing frequency per day of compartment access revealed no significant differences between bold and shy cats using either method of defining this behavioural style (Table 5.4).

When the observer rating system was used to classify the cats into bold/shy, Mann-Whitney U tests comparing compartment use revealed that shy cats spent a significantly greater percentage of time in BOX than did bold cats. None of the other compartments showed a difference in percentage of time spent using this method of classification, and no significant difference was found in percentage of time spent in any compartment using the emergence test to dichotomise bold/shy individuals (Table 5.4).

Table 5.4 Median, IQR, and Mann-Whitney U tests comparing bold and shy cats of frequency of compartment visit, and percentage of time spent in each compartment for bold and shy cats (as classified by both observer rating and the emergence test)

			Observer Rating			Emergence Test		
Parameter	Compartment	Bold/Shy	Median	IQR	Mann-Whitney U P-value	Median	IQR	Mann-Whitney U P-value
Frequency of visit	CTRL	Bold	4	4	0.44	2	4	0.09
		Shy	2	3		5	4	
	TOY	Bold	4	4	0.35	2	4	0.27
		Shy	2	4		4	6	
	SHELF	Bold	4	4	0.80	3	4	0.70
		Shy	3	6		4	6	
	BOX	Bold	4	5	0.23	5	4	0.66
		Shy	7	4		6	5	
Percentage of time	CTRL	Bold	4	7	0.57	1	6	0.14
		Shy	3	3		3	7	
	TOY	Bold	2	21	0.32	2	4	0.19
		Shy	2	6		4	18	
	SHELF	Bold	6	32	0.60	6	48	0.94
		Shy	6	17		10	29	
	BOX	Bold	37	53	0.04	53	47	0.90
		Shy	66	29		58	51	

5.5 Discussion

Choice experiments can provide valuable information about an animal's behavioural needs, especially when more complicated motivation tests are impractical or too costly. Additionally, Kirkden and Pajor (2006) point out that if the alternatives the animal is choosing between – even if they are 'non-substitutes' (i.e. alternative ways of satisfying the same positive motivation) – inform the resources provided by an animal care giver, a choice test may still be a reasonable method of assessment. In this study, training the cats to use the cat flap doors proved quite difficult and the speed at which cats learned to use them was quite varied – likely due to cat-level characteristics such as previous experiences (i.e. if they have used a cat flap door before), body weight (i.e. heavier cats may find it easier to push the doors open), or behavioural styles (i.e. bold cats may be more likely to engage with the novel door than shy cats). Training time needed to be short, as it was desirable to investigate the visits made as early after arrival at the facility as possible. Each choice chamber was fitted with two doors accessing each external compartment, with the intent being to have one act as an entrance (which could be weighted) and one to act as an exit (which would not be weighted). However, training the cats to simply understand how to use the doors when they could swing freely – allowing access both in and out – took more time than expected. Therefore, the additional time required for the cats to then learn which door was the entrance and which door was the exit would have meant that the actual experiment would begin even longer after arrival at the facility. For this reason, a simple choice experiment was deemed the best option. As the data on compartment usage was assessed using Onset HOBO U11 3 State/1 Event Data Loggers and not direct observation, it is possible that the recording method may have influenced the agreement or reliability of the data. However, as outlined in Ch 6 (this text) these parameters surpassed the established minimum levels agreement and reliability for all external compartments (TOY, SHELF, BOX, and CTRL).

As the emergence test showed poor agreement with the observer rating in this study, classification of a cat as either bold or shy in this study was unreliable. As a result, behavioural style was not included as a predictor in the linear model, as it was unclear whether to include the bold/shy classifications based on the results of the behavioural test of the observer ratings. Therefore, a comparison of compartment use of bold and shy cats was conducted using the dichotomous classifications produced by both systems. The emergence test requires further validation with a larger sample size to evaluate its potential to discriminate between bold and shy cats.

The analysis revealed no effect of days since considered reliably trained, and thus all further analyses were conducted on medians of daily data, however, the lack of an effect of time could have been influenced by the amount of time it took to train individuals. Cats arrived at the facility and were placed into their enclosures with no flaps in the cat doors. The next day the flaps were placed in the doors and training commenced. Cats continued with training daily every day until the investigator was certain – through visual observation or reviewing activity on the data loggers – that the cat could use the doors at will. The first day of the experiment (i.e. the day on which recordings were used in the subsequent analysis) was considered as the first full calendar day without training. In effect, this meant that the shortest time between arrival at the facility and beginning of experiment was two days. In reality, it was often longer than that (range 2-8 days). McCune (1992) found that cats exhibited the greatest signs of stress due to confinement in the first 24h while both Smith et al. (1994) and Kessler and Turner (1997) found that most behavioural signs of stress stabilised after the first four days. It is possible that in the current study there was a change in visit frequency/time allocation over time, but that the experimental design was not sensitive enough to capture these data during the pre-habituation phase. If cats interact with resources differently during the pre-habituation phase, it may be that shelters should prioritise providing one type of enrichment before habituation, and another type afterwards. Further investigation of comparison of visit frequencies/time allocation exhibited before and after habituation is necessary to substantiate this speculation.

No difference was found in individual compartment use between dark and light periods for the percentage of time spent. This is not surprising as the percentage of time spent active in laboratory cats has been found to peak around the twilight hours from behavioural data (Kavanau, 1971) and electroencephalographic analysis of brain activity (Sterman et al., 1965). However, in the current study frequency of visit to each compartment was significantly different or close to significantly different between dark and light periods for all compartments. Both studies mentioned above only presented data on activity as a percentage, not on frequency of activity across time. The data presented in Figure 5.2 of the current study demonstrate that while the percentage of activity is not different between the day and night periods, the frequency of active bouts may be greater in the light periods, as was found by Sterman et al. (1965)

Cats spent a significantly greater percentage of time in BOX than in CTRL or TOY. The parameter frequency of compartment visit showed a similar pattern, but could not reach statistical significance in the pairwise comparisons, possibly due to the relatively high degree of individual variation compared to the magnitude of the median. Hiding has been described as an important behaviour for cats exposed to potentially aversive conditions. Rochlitz et al. (1998) found that cats housed in a quarantine facility spent significantly less time hiding after the first month, and Carlstead et al. (1993b) found that the percentage of time spent hiding increased in zoo-housed leopards for the first week following translocation to a novel, barren environment. These studies suggest hiding behaviour may represent a key strategy of captive felids for coping with novel – and presumably stressful – circumstances. Indeed, Carlstead et al. (1993a) found that cats exposed to a presumed stressor (unpredictable handling and routine) exhibited more hiding behaviour than did those in the control group, and that hiding behaviour was negatively correlated with cortisol concentration, further supporting the role of this behaviour in coping with stressful conditions. Kry and Casey (2007) found that cats provided with a hiding box registered greater daily reductions in Cat-Stress-Scores on each day assessed than did the control group, performed more “true resting behaviour” and approached an observer significantly more often in an approach test with the kennel door closed, while not reducing their likelihood of adoption as a result of concealment. Each of these studies concluded with the recommendation of providing hiding opportunities to confined cats. Additionally, Griffith et al. (2000) found that the effects of a feline facial pheromone being investigated for its anxiolytic properties were increased by providing it in combination with a hiding opportunity in cats housed in a veterinary hospital.

The fact that the hiding compartment was used so significantly over the others raises the question whether it is environmental enrichment at all. While the Ch 1 of this thesis has defined environmental enrichment as “any addition to the environment of an animal resulting in a presumed increase in the environment’s quality, and a subsequent presumed improvement to the animal’s welfare”, some authors (i.e. Duncan and Olsson, 2001) contend that if a proposed environmental enrichment item alleviates a negative state rather than promoting a positive one, it should not be termed environmental enrichment, but rather an environmental requirement. These authors may claim that the much greater allocation of time to the hiding box may be evidence that not providing an opportunity for these individuals to hide is failing to cater to their basic needs. Under this theoretical framework, provisions that alleviate impoverishment would be recommended before any provisions that promote a positive state. While theory may differ between this definition and the one used in this thesis, the practical outcomes are the same as under the definition used in this thesis, i.e. evidence that certain types of environmental enrichment which are used demonstrably more by subjects or provide some type of benefit (i.e. reduced faecal glucocorticoid concentrations or cat-stress-scores) would result in its provision. SHELF was not found to be significantly different from any of the other compartments using any of the parameters after Bonferroni-adjustments, although the comparison between BOX and SHELF approached significance for percentage of time spent in each compartment. This suggests that the resource in this compartment may have been the second most used enrichment item. This suggestion is supported by the fact that the SHELF has the second highest median value (Table 5.1) for each parameter. However, the magnitude of SHELF's median is much closer to that of CTRL and TOY than it is to that of BOX in all of the parameters, and the relatively high degree of individual variation is likely responsible for the lack of differentiation from BOX. Therefore, while there is some evidence that the resource in SHELF may be the second most used enrichment item, this may have been highly dependent on some individual characteristic. Additionally, this resource may be of substantially more importance in communally housed cat environments, where perching for vigilance may be of greater biological relevance.

Since at least one of the enrichment compartments was found to be used more than the empty compartment, this suggests that cats allocate more of their time in the presence of enrichment over its absence. However, this relationship is dependent upon the type of enrichment provided, as is suggested by the finding that not all enrichment compartments were used more than the empty compartment.

Convincing evidence was presented above that the subjects of this study consistently allocated more of their time to the BOX compartment, however, if different behaviours are associated with the various environmental enrichment items present in each of the compartments, and these behaviours take different amounts of time to exhibit, then perhaps what could be interpreted as a preference is simply a reflection of the activity budget of the subjects, and has little to do with the animals choosing environments that will improve their welfare. For example, an animal is likely to spend a large amount of its time sleeping and a very small amount of its time drinking, but this does not mean that drinking is not important to the welfare of the animal. Likewise, in the present study, if a cat always sleeps in BOX, and always plays in TOY, it is likely that by reflection of its activity budget it is bound to spend more time in BOX, even if it does derive benefit from playing in TOY (perhaps by alleviating boredom from lack of stimulation or frustration from not being able to perform prey-catching behaviours). Further investigation of the importance of this time allocation would have been possible in a consumer-demand test, but unfortunately due to the reasons given above, this was not possible in the current study. Therefore, future investigation of the biological or behavioural effects of the provision of some or all of the different environmental enrichment items offered here is necessary to draw conclusions about the nature of this time allocation (discussed further below).

In the analysis of cat-level predictors, separate for each compartment and each parameter, surrendered cats were found to spend a significantly higher percentage of their time in CTRL than did stray cats. It is possible that stray cats were used to the dangers of being out in the open and thus avoided large empty spaces, or were used to having to accumulate their own resources – unlike surrendered cats who would have been provided for by their owners – and thus stayed in the vicinity of the other, resource-laden compartments. However, this relationship was not expected and further research is necessary to substantiate these possible interpretations.

For SHELF, the only cat-level predictor found to be significant was in which choice chamber they were housed. Cats housed in choice chamber 3 spent a significantly higher percentage of their time in SHELF than did cats housed in choice chamber 2. Figure 5.1 illustrates that choice chamber 1 was located in a room by itself, while choice chambers 2 and 3 were housed in one room together. This was so that one of the rooms would have enough space in which to conduct animal husbandry procedures. As SHELF was closest to the other choice chamber in the room for cats in choice chamber 3, it is possible that the comparatively increased use of this compartment by these cats was influenced by their proximity to other cats. In this scenario, the cats in choice chamber 3 allocated more time to the SHELF compartment in order to be closer to the cats in choice chamber 2. This scenario is supported by the fact that cats in choice chamber 2 spent most of their time in the compartment closest to choice chamber 3. However, as the compartment closest to choice chamber 3 was BOX – the compartment the cats used most anyway – this evidence could be misleading. Furthermore, there was no significant difference in percentage of time spent in any of the compartments between cats housed in choice chamber 1 or 2, as would be expected if the need to be closer to other cats was the driving force between the relatively increased time spent in SHELF by cats in choice chamber 3. Another possibility is explained by each chamber's placement within the room, with respect to other structures in the room. Figure 5.1. shows that cats in choice chambers 1 and 2 could have human activity occur, or their enclosure approached, from several angles while visible human activity or approach could only on one side for choice chamber 3 – almost directly in front of SHELF. This might mean that cats in choice chamber 3 might have associated the perching opportunity in SHELF with human contact, and might have spent more time waiting in that compartment for the opportunity for social enrichment in the form of human interaction.

Additionally, unlike the other choice chambers, this one was tightly surrounded by walls on three sides, and the side without walls was most closely associated with SHELF. For this reason, perhaps cats in choice chamber 3 spent relatively more time in SHELF because they saw it as their greatest chance of escape. The most likely reason cats in choice chamber 3 used SHELF relatively more than did cats in the other choice chambers would require further investigation. However, this result illustrates that choice tests can be highly sensitive to experimental conditions (e.g. ambient temperature; Fraser, 1985).

The comparison of compartment use between bold and shy cats (using both systems of classification) showed the only significant difference to be for BOX. Using the observer rating classification system, BOX was found to be used for a larger percentage of overall time by shy cats than bold cats. It makes biological sense that shy cats would be more likely to seek the protection of a hiding position than would bold cats, and thus there is some support for the hypothesis that shy cats use hiding to cope with stress of exposure to unfamiliar people, animals, and environments. While this hypothesis is supported by the results of two parameters, it is worth noting that as the opinion method of assessing cats to be bold or shy relies on subjective observation over time; it is possible that the observed time spent in the hiding box could have influenced the assessment. This means that these two variables may not be entirely independent, although this was impossible to avoid given the experimental design. Additionally, the same evidence was not produced when classifying cats according to the emergence test, suggesting that either more refinement of this dichotomous behavioural test may be necessary, or that shy cats do not actually consistently use the hiding box significantly more than bold cats.

The fact that these analyses did not reveal any differences in the use of TOY between bold and shy cats, means that contrary to the hypothesis that bold cats would be more likely to actively engage with prey-simulating toys in order to stave off the boredom that can be associated with lack of physical complexity in otherwise barren shelter cages was not met. However, there is some evidence that while the prey-simulating feature of a toy may make it the most appealing to a cat, novelty plays a key role in maintaining their interest. de Monte and Le Pape (1997) exposed cats to two objects: a stationary log which elicited rubbing and paddling behaviours, and a mobile ball which stimulated play and simulated prey. They found that cats interacted significantly more with the prey-simulating ball, however with both items a pronounced habituation effect was observed; the second day of exposure to the objects was accompanied with a significant drop in interaction time. Similarly, Hall et al. (2002) found that three sequential interactions with a toy were enough to cause almost complete habituation. However, they found that changing sensory characteristics of the toy elicited significant disinhibition, and renewed interest in the toy. They concluded that while object play is largely elicited by prey-like characteristics, it relies on change or novelty to be maintained. Since this habituation seems to happen so quickly, it is possible that cats did allocate more time to the prey-simulating toy in the current study, but that habituation occurred between arrival at the facility and the start of the actual experiment. Perhaps if the toy had been changed daily, the compartment would have generated more interest and potentially revealed a difference between bold and shy cats. However, as the practical goal of this study was to advise shelter staff how best to enrich cats, perhaps daily replacement of appropriate toys may not be practical from a sterilisation or labour standpoint.

Despite large variation in compartment usage, few cat-level characteristics found to be significant predictors. This suggests that there may be one or more cat-level characteristics that were not considered and contributed greatly to the movement and time distribution of the cats in this study. Potential candidates for such characteristics include previous experiences and/or some sort of genetic component. Unfortunately many of these cat-level characteristics are hard to quantify, or impossible to ascertain from shelter cats.

In previous studies, cats have shown preferences for different physical properties of enrichment items. Smith et al. (1994) provided cats with a raised platform with four equally sized squares of wood, plastic, metal, and fabric substrate. Cats overwhelmingly positioned themselves on the wooden square. However, seemingly no investigation of the choices made for, or motivation to access, different types of environmental enrichment have been conducted for this species. In farmed mink, Mason et al. (2001) showed that despite increasing 'costs', subjects continued to work to access a pool of water in which they could swim, and excreted elevated concentrations of cortisol when access to the swimming compartment was barred. This suggests that they are highly motivated to perform this behaviour, and that the inability to engage in it may contribute to a negative state. In laboratory rabbits, Seaman et al. (2008) showed a preference and motivation to be in a compartment containing a platform, but not to actually use it. The authors suggested that the platform may have represented a potential object to hide under when threatened, and thus proximity to it increased their feeling of safety. However, the rabbits showed an even stronger motivation for access to conspecifics, suggesting that access to intra-species social environmental enrichment may most substantially reduce a negative state or increase a positive one. For some species, inter-species social environmental enrichment via human interaction has been suggested as preferred (Deluca and Kranda, 1992) or more important to an animal's well-being (Wells, 2004). In the current study, the fact that the cats in choice chamber 3 were more likely to spend time in the compartment in the direction from which all human approaches took place may be evidence of the cats exhibiting a desire for human interaction. There is some support for this in a study conducted by Denenberg (2003), which investigated the preferences of cats for different toys; the cats tended to prefer the toys that involved human interaction. While the author interpreted these findings as preference for toys that stimulated predation or feeding, it is also possible that the cats preferred these toys because of the added benefit of interaction with humans.

5.5.1 Conclusions

These results indicated that the cats in this study consistently allocated more of their time to being in the proximity of some types of enrichment over others, and there was some indication that cats were more active during the light period than during the dark period. Overall the cats allocated more of their time in proximity of the hiding opportunity. There was some evidence that the shelf compartment may have been the enriched environment in which the cats allocated the second most of their time, however this may be highly dependent on some individual variation characteristic. When analysing each compartment separately, surrendered cats were found to spend a greater percentage of time in CTRL than did strays. Percentage of time spent in TOY was not influenced by any cat-level predictors. Cats housed in choice chamber 3 were found to spend a greater percentage of time in the SHELF than the cats housed in choice chamber 2, highlighting how choice tests are highly sensitive to experimental conditions. Shy cats were found to spend a greater percentage of time spent in BOX than did bold – but only using one of the two methods of classification, suggesting unreliability with either the results or one of the tests. This is supported by the fact that the agreement between the results of the emergence test and the observer ratings failed to reach minimum acceptability. These findings suggest that resource utilization is relatively uniform across cats – although certain cat-level characteristics may have the potential to influence this – and that hiding may satisfy a basic need for cats in caged conditions.

5.5.2 Further work

The results of this study raised some interesting questions. Due to the study design, differential expression of resource use before and after habituation may have been lost. Since the pre-habituation period has been isolated as the most stressful period (Kessler and Turner, 1997; McCune, 1992; Smith et al., 1994), investigation into resource use during this period may be very helpful. Some cat-level predictors were found to influence compartment use in unexpected ways. Investigation into why stray /surrendered cats spent different percentages of their time CTRL may reveal behavioural strategies developed in varying conditions. Finally, with enrichment usage established, investigation of the effect of providing or withholding these forms of enrichment may prove valuable (Fraser and Nicol, 2011), and even solve some of the problems created by conducting a simple choice test and not the more informative consumer demand theory test. If providing cats with the type of enrichment they allocate more of their time in proximity to affects behavioural and physiological measurements considered to be relevant to welfare, then this would provide additional support for providing their these enrichment items.

5.6 References

Barnett, L.A., 1984. Research note: Young children's resolution of distress through play. Journal of Child Psychology and Psychiatry 25, 477-483.

Blom, H.J., Van Vorstenbosch, C.J., Baumans, V., Hoogervorst, M.J., 1992. Description and validation of a preference test system to evaluate housing conditions for laboratory mice. Appl. Anim. Behav. Sci. 35, 67-82.

Carlstead, K., Brown, J.L., Strawn, W., 1993a. Behavioral and Physiological Correlates of Stress in Laboratory Cats. Appl. Anim. Behav. Sci. 38, 143-158.

Carlstead, K., Seidensticker, J., Brown, J.L., 1993b. Behavioral and adrenocortical responses to environmental changes in leopard cats (*Felis bengalensis*). Zoo Biol. 12, 321-331.

Dawkins, M.S., 1983. Battery hens name their price: Consumer demand theory and the measurement of ethological ‘needs’. Anim. Behav. 31, 1195-1205.

Deluca, A.M., Kranda, K.C., 1992. Environmental enrichment in a large animal facility. Lab Anim. 21, 38-44.

deMonte, M., LePape, G., 1997. Behavioural effects of cage enrichment in single-caged adult cats. Anim. Welfare 6, 53-66.

Denenberg, S., 2003. Cat toy play trial: A comparison of different toys. Proceedings of the Annual Scientific Symposium of Animal Behaviour, American Veterinary Society of Animal Behaviour, Denver Colorado.

Duncan, I., Olsson, I., 2001. Environmental enrichment: from flawed concept to pseudo-science. International Society of Applied Ethology 35th International Congress, 73.

Duncan, I.J.H., 2005. Science-based assessment of animal welfare: farm animals. Revue Scientifique et Technique - Office International des Épizooties 24, 483-492.

Duncan, I.J.H., 1992. Measuring preferences and the strength of preferences. Poult. Sci. 71, 658-663.

Ellis, S.L.H., 2009. Environmental enrichment: practical strategies for improving feline welfare. J. Feline Med. Surg. 11, 901-912.

Fraser, D., 1985. Selection of bedded and unbedded areas by pigs in relation to environmental temperature and behaviour. Appl. Anim. Behav. Sci. 14, 117-126.

Fraser, D., Milligan, B.N., Pajor, E.A., Weary, D.M., 1997. A scientific conception of animal welfare that reflects ethical concerns. Anim. Welfare 6, 187-205.

Fraser, D., Nicol, C.J., 2011. Preference and motivation research, in: Appleby, M.C., Mench, J.A., Olsson, I.A.S., Hughes, B.O. (Eds.), Animal Welfare. CABI, Wallingford, pp. 183-199.

Griffith, C.A., Steigerwald, E.S., Buffington, C.A.T., 2000. Effects of a synthetic facial pheromone on behavior of cats. J. Am. Vet. Med. Assoc. 217, 1154-1156.

Hall, S.L., Bradshaw, J.W.S., Robinson, I.H., 2002. Object play in adult domestic cats: the roles of habituation and disinhibition. Appl. Anim. Behav. Sci. 79, 263-271.

Kavanau, J.L., 1971. Locomotion and activity phasing of some medium-sized mammals. J. Mammal. , 386-403.

Kessler, M.R., Turner, D.C., 1997. Stress and adaptation of cats (*Felis silvestris catus*) housed singly, in pairs and in groups in boarding catteries. Anim. Welfare 6, 243-254.

Kirkden, R.D., Pajor, E.A., 2006. Using preference, motivation and aversion tests to ask scientific questions about animals' feelings. Appl. Anim. Behav. Sci. 100, 29-47.

Kry, K., Casey, R., 2007. The effect of hiding enrichment on stress levels and behaviour of domestic cats (*Felis sylvestris catus*) in a shelter setting and the implications for adoption potential. Anim. Welfare 16, 375-383.

Landis, J.R., Koch, G.G., 1977. The Measurement of Observer Agreement for Categorical Data. Biometrics 33, pp. 159-174.

Mason, G., Cooper, J., Clarebrough, C., 2001. Frustrations of fur-farmed mink. Nature 410, 35-36.

McCune, S., 1992. Temperament and the welfare of caged cats. Ph.D. Thesis, University of Cambridge.

Podberscek, A.L., Blackshaw, J.K., Beattie, A.W., 1991. The Behavior of Laboratory Colony Cats and their Reactions to a Familiar and Unfamiliar Person. Appl. Anim. Behav. Sci. 31, 119-130.

Rochlitz, I., 1999. Recommendations for the housing of cats in the home, in catteries and animal shelters, in laboratories and in veterinary surgeries. J. Feline Med. Surg. 1, 181-91.

Rochlitz, I., 2000. Recommendations for the housing and care of domestic cats in laboratories. Lab. Anim. 34, 1-9.

Rochlitz, I., 2005. A review of the housing requirements of domestic cats (*Felis silvestris catus*) kept in the home. Appl. Anim. Behav. Sci. 93, 97-109.

Rochlitz, I., Podberscek, A., Broom, D., 1998. Welfare of cats in a quarantine cattery. Vet. Rec. 143, 35-39.

Seaman, S.C., Waran, N.K., Mason, G., Richard B. D'Eath, 2008. Animal economics: assessing the motivation of female laboratory rabbits to reach a platform, social contact and food. Anim. Behav. 75, 31-42.

Shepherdson, D.J., Carlstead, K., Mellen, J.D., Seidensticker, J., 1993. The Influence of Food Presentation on the Behavior of Small Cats in Confined Environments. Zoo Biol. 12, 203-216.

Smith, D.F., Durman, K.J., Roy, D.B., Bradshaw, J.W., 1994. Behavioural aspects of the welfare of rescued cats. The Journal of the Feline Advisory Bureau 31, 25-28.

Spinka, M., Newberry, R.C., Bekoff, M., 2001. Mammalian play: training for the unexpected. Q. Rev. Biol. , 141-168.

Sterman, M.B., Knauss, T., Lehmann, D., C.D Clemente, 1965. Circadian sleep and waking patterns in the laboratory cat. Electroencephalogr. Clin. Neurophysiol. 19, 509-517.

Wells, D.L., 2004. A review of environmental enrichment for kennelled dogs, *Canis familiaris*. Appl. Anim. Behav. Sci. 85, 307-317.

Wemelsfelder, F., 1997. Life in captivity: its lack of opportunities for variable behaviour. Appl. Anim. Behav. Sci. 54, 67-70.

6 ASSESSING THE AGREEMENT AND RELIABILITY OF AN AUTOMATED METHOD (HOBO DATA LOGGER) TO RECORD THE MOVEMENT OF A CAT IN AND OUT OF CHAMBERS

6.1 Abstract

Behavioural observation of videotapes can be time consuming. Remote monitoring technology exists which can alleviate the need to watch videos to quantify certain types of behaviours. In order to validate such methods, their agreement and reliability must be evaluated to ensure the data generated are of sufficient quality. In this study, the agreement and reliability of an Onset HOBO U11 3 State/1 Event Data Logger - U11-001 – and related data cleaning method – was compared against video observation of the same period, for its ability to record the frequency of compartment visit and percentage of time spent in each compartment. For 10 days, six cats were housed in one of three plus-shaped choice chambers offering four different types of environmental enrichment (EE) in the external compartments, and food and litter in the central compartment. Continuous behavioural observation of two 12-h periods of video data from each cat were compared against the data-logger results from the same period. When looking at each compartment separately and at each 12-h period separately (cumulative data from all EE compartments), frequency of compartment visit and percentage of time spent in each compartment had acceptable agreement (CCC≥0.60) and reliability (ICC≥0.60). These findings validate the use of data loggers and the data cleaning method to monitor cat behaviour in the EE compartments of these plus-shaped choice chambers.

6.2 Introduction

Behavioural studies are often evaluated via observation of videotapes, unfortunately watching tapes can be time consuming and tedious. Blom et al. (1996) employed and validated an elaborate LED sensor system to remotely monitor movement of animals between cages. However, the technology developed was complicated and difficult to replicate. Onset HOBO data loggers, through the use of a simple reed and magnet switch, have the potential to electronically monitor the opening and closing of cat flap doors in a plus-shaped choice test, and thus to monitor the movement of a cat between chambers. However, due to incidents of the cats opening the doors but not going through them, it is possible that the raw data produced by the loggers would not give a true picture of the cats' entries into the chambers. A method of cleaning the raw data produced by the loggers could dampen this effect greatly. Therefore, the objective of this study was to develop a data cleaning method and evaluate if an Onset HOBO data logger – after application of a data cleaning method – can be reliably used to assess movement of cats in and out of a choice chamber, when compared to analysis of video data of the same period.

6.3 Methods

6.3.1 Subjects and housing

This study was conducted on a subset of the cats in Ch 5 (this text): six domestic cats obtained from a local shelter, of which one was an intact male, and five were females of unknown reproductive status. The mean length of stay in the shelter before transport approximately 5 km to the study site was 4 ± 1.1 days (±SD), and mean estimated age of the cats included in this study was 2.5 ± 1.00 years. Three of the cats had been surrendered by their owner and three of them were strays. Housing and care was identical to the conditions described in Ch 5.

6.3.2 Procedure and apparatus

A detailed description of the procedure and apparatus is provided in Ch 5. In brief, cats were housed singly in one of three plus-shaped 'choice chambers' with five separate compartments. The centre compartment (HOME) contained food and water dishes, and each external compartment contained a different type of environmental enrichment: compartment 1 (CTRL) contained no additional enrichment items, compartment 2 (TOY) contained a prey-simulating toy, compartment 3 (SHELF) contained an elevated shelving unit, and compartment 4 (BOX) contained a hiding box.

6.3.3 Data collection and cleaning

HOBO data

Each choice chamber had a HOBO data logger with four lines monitoring the cat flap doors, with one line dedicated to each compartment. Lines 1-3 of the HOBOs were the 'state' lines, and recorded a '0' every time a door opened and a '1' every time a door closed. Line 4 was the 'event' line, and only registered a data point when doors opened. In theory, researchers should simply be able to look at the '1' data points, interpreting the first opening as a cat entering the compartment, and the second as the cat leaving the compartment (at least for lines 1-3). However, there were at least five different issues that complicated the data. First, cats had a tendency to rub up against the doors, which would open the door enough to trigger the reed switch, without the cat changing compartments. Second, sometimes the door would swing shut with enough force that it would swing back and forth a few times before finally coming to a stop. The data loggers would record each of these instances as a separate data point, even though the cat had only gone through the door once. Third, sometimes it would take a cat more than one attempt to get though a door, meaning it would nudge the door with its paw or face a few times before successfully changing compartments. This too would register as more than one data point. Fourth, it was possible that while performing regular cleaning or experimental procedures, an animal care technician or the researcher could inadvertently push one of the doors, registering a false change in compartment. Finally, on at least one instance, a cat was observed to sit in the centre compartment and simply bat the door open repetitively. Due to these complications of the collection method, some data cleaning was required.

Cleaning the HOBO data

Different methods were investigated – including eliminating all data points that were ≤ 10 s, or times when the door was only open for ≤10 s. But considering that a successful compartment visit must have at least two door openings (an entrance and an exit) and these techniques often left odd numbers of data points attributed to compartments, the results were uninterpretable. The most successful data cleaning technique employed involved first eliminating all data points registering door close events (as only a door opening *or* a door closing is necessary to indicate a cat changing compartments). Next, any instances of a door opening only once before another compartment was accessed were eliminated; since a door must open twice for a cat to enter and exit a compartment, a singular access point would represent an instance where a door opened, but no compartment change took place. Furthermore, in cases where a door opened more than twice in a row, all but the first and last data points were eliminated in order to reduce the effect of instances where a door was simply nudged.

The data that remained revealed unambiguously which compartment the cat was in at all times. Frequency of compartment visit and percentage of time spent in each compartment were calculated for each day, and medians of these daily values were used for subsequent analyses.

Video data

For the purposes of validating the cleaned HOBO data, two 12-h periods of video data from each cat were analysed. All sample periods took place between 0600 and 1800 h, and no two sample periods took place on the same day. The breakdown of which cats were sampled on which days is presented in Table 6.1. Recordings from videotapes were observed and each change of compartment was recorded via continuous observation, resulting in a data set reporting frequency of compartment visit and percentage of time spent in each compartment. This analysis was conducted using Noldus Observer 5 (Noldus Information Technology, Wageningen, The Netherlands).

Table 6.1 Breakdown of cat IDs and sample dates for videotapes analysed for validation of cleaned data logger data

Observation #	Cat ID	Choice chamber #	Sample Date (DD/MM/YY)
1	1	1	040311
2	1	1	060311
3	2	2	080311
4	2	2	090311
5	3	3	050311
6	3	3	070311
7	4	1	160311
8	4	1	210311
9	5	2	190311
10	5	2	230311
11	6	3	200311
12	6	3	240311

Note: for visual representation of where choice chambers 1-3 are located in relation to each other see Table 5.1, Ch5.

6.3.4 Statistics

Agreement and reliability between the HOBO data and the video data were first assessed for each compartment separately across all observations (n=12) to see if there was a systematic bias present in any one of the compartments. Next, agreement and reliability were assessed across compartments (n=4) within each of the 12 individual observations, and the median for each parameter was calculated for use as the overall result for each test.

Agreement was assessed – both by compartment and by observation – using the concordance correlation coefficient (CCC) (Lin, 1989) for the variables frequency of compartment visit and percentage of time spent in each compartment. Reliability was assessed – both by compartment and by observation – using the intraclass correlation coefficient (ICC) (Lessells and Boag, 1987). Interpretation of agreement and reliability parameters followed Martin and Bateson's (2007) guide to interpreting correlation coefficients for CCC and ICC. Descriptive statistics and ICCs were calculated in Minitab® 15 statistical software, while CCCs were conducted using Stata 10 statistical software.

6.3.5 Ethical approval

This project was approved by UPEI's Animal Care Committee under protocol number 09-051, and followed the guidelines of the Canadian Council of Animal Care's "Guide to the Care and Use of Experimental Animals".

6.4 Results

Overall median values were calculated for frequency of compartment visit and percentage of time spent in each compartment. These data have been presented by compartment and by observation, in Tables 6.2 and 6.3 respectively.

Table 6.2 Median values for frequency of compartment visit and percentage of time spent in each compartment, presented by compartment (n=12)

	Frequency of compartment visit		Percentage of time spent in each compartment	
Compartment	Video	HOBO	Video	HOBO
HOME	16.0	14.0	4.2	11.6
CTRL	6.0	3.5	5.5	3.2
TOY	5.0	4.0	6.2	4.8
SHELF	4.5	2.0	18.2	6.6
BOX	4.5	4.0	49.2	34.4

Table 6.3 Total number of visits to all compartments and compartment in which the highest percentage of time was spent (followed by the percentage), presented by observation (compartment n=5: 4 enriched compartments and the centre compartment)

	Total number of visits		Compartment in which greatest % of time was spent	
Observation	Video	HOBO	Video	HOBO
1	43	37	SHELF (54%)	SHELF (53%)
2	17	11	BOX (78%)	BOX (69%)
3	31	20	TOY (85%)	TOY (86%)
4	39	37	BOX (67%)	BOX (69%)
5	33	19	BOX (84%)	BOX (85%)
6	27	26	BOX (61%)	BOX (37%)
7	33	37	SHELF (55%)	SHELF (53%)
8	65	42	BOX (69%)	BOX (70%)
9	11	7	SHELF (96%)	HOME (57%)
10	33	31	BOX (87%)	BOX (87%)
11	73	57	TOY (39%)	BOX (32%)
12	15	5	SHELF (85%)	SHELF (85%)

6.4.1 By compartment

The CCC conducted by compartment revealed that frequency of visit achieved high agreement in all compartments except SHELF, where the agreement was moderate. For percentage of time spent, all compartments showed high to very high agreement, except HOME, which registered a negative correlation, indicating no agreement (Table 6.4).

The ICC conducted by compartment revealed that frequency of visit achieved high reliability in all compartments except SHELF, which reached moderate reliability. For percentage of time spent, all compartments achieved high to very high reliability in all compartments except HOME, which registered a negative correlation, indicating no reliability (Table 6.4).

Table 6.4 CCC and ICC for analyses conducted within compartments, across observations (n=12)

	Frequency of compartment visit		Percentage of time spent in each compartment	
Compartment	CCC	ICC	CCC	ICC
HOME	0.82	0.83	-0.06	-0.38
CTRL	0.72	0.73	0.95	0.95
TOY	0.86	0.86	0.95	0.95
SHELF	0.60	0.60	0.86	0.87
BOX	0.87	0.88	0.95	0.95

Although not all compartments reached the established acceptable levels of agreement or reliability in each measure, most reached at least moderate levels. The main exception to this was HOME, which failed to achieve acceptable levels of both agreement and reliability for percentage of time spent in each compartment, reaching only slight correlation of the parameters. This suggested a systematic bias in this compartment, and it was therefore dropped from further analyses.

6.4.2 By observation

The median CCC for analyses conducted across all compartments except HOME within individual observations revealed high agreement for frequency of compartment visit (8/12 observation had greater than or equal to high agreement), and very high agreement for percentage of time spent in each compartment (10/12 observation had greater than or equal to high agreement) (Table 6.5).

Table 6.5 CCC and ICC for analyses conducted within observations, across compartments (without HOME) (n=4)

	Frequency of compartment visit		Percentage of time spent in each compartment	
Observation	CCC	ICC	CCC	ICC
1	0.70	0.74	0.99	0.92
2	0.85	0.88	0.98	0.98
3	0.73	0.77	1.00	1.00
4	0.66	0.73	0.90	0.97
5	0.67	0.71	1.00	1.00
6	0.84	0.88	0.80	0.84
7	0.98	0.99	0.84	0.87
8	0.39	0.36	1.00	1.00
9	0.93	0.73	0.67	0.73
10	0.97	0.98	1.00	1.00
11	0.88	0.34	0.55	0.62
12	0.28	0.49	1.00	1.00
Overall median (IQR)	**0.79 (0.27)**	**0.73 (0.29)**	**0.99 (0.19)**	**0.98 (0.15)**

The median ICC for analyses conducted across all compartments except HOME within individual observations revealed high reliability for frequency of compartment visit (9/12 observation had greater than or equal to high agreement), and very high reliability for percentage of time spent in each compartment (11/12 observation had greater than or equal to high agreement) (Table 6.5).

6.5 Discussion

Agreement and reliability analyses fairly consistently revealed acceptable values for frequency of compartment visit and percentage of time spent in each compartment for all compartments except HOME, suggesting a systematic bias in that compartment, and resulting in its removal from subsequent analyses. The median agreement and reliability figures taken from individual observations revealed high correlation for frequency of compartment visit, and very high correlation for percentage of time spent in each compartment. This validates the use of data loggers and the data cleaning method to monitor cat behaviour in these plus-shaped choice chambers.

6.5.1 The bias in HOME

The systematic bias in HOME is likely a product of the data cleaning method. Although this method has shown a high degree of validity, there are still some issues that are not resolved. While this procedure eliminated the problem of odd numbers of door opening events rendering data uninterpretable, there were two problems with the data that this method did not fix. First, it did not take into account a situation where an animal continually moved between the centre and a specific external compartment, as it would eliminate all data points in a sequence, except the first and last, treating several visits in a row as one long visit, reducing the amount of time spent in the centre compartment (HOME). Second, it would treat accidental knocks of one of the doors by an animal care technician or the researcher as if they were evidence that the cat was in (or had re-entered) HOME and performed an action such as rubbing up against the door – therefore falsely augmenting the time recorded as spent in HOME. There is room for error with the cleaning method, and since all external compartments connect only with HOME, this means that this compartment is subject to the residual error from all compartments, rendering its results least reliable and with poor agreement against video data.

6.5.2 Choice of agreement and reliability parameters

The choice of parameters to measure agreement and reliability was difficult. The CCC was the chosen method of agreement because it is the correlation between the two readings that fall on the 45^0 line (Lin, 1989), but also because upon investigation of the Bland-Altman method (Bland and Altman, 1986) the differences were not found to be normally distributed. There is one potential problem with using the CCC to investigate within observations. The use of each compartment would not be independent of each other, as using one compartment could take away from time using another compartment. Such a dependence is most clear when considering the percentage of time spent in the various compartments. However, the removal of HOME has drastically reduced the inter-dependence of compartment use, rendering this parameter more applicable.

The ICC is the most common method of assessing reliability of behavioural data, as it takes into account variation over time of each individual. Unfortunately, as the calculation works within an ANOVA framework, there are several assumptions implicit in using this statistic. As many of the outcome variables being analysed here cannot be transformed to comply with these assumptions, the results must be interpreted with caution. Many authors have argued that the results of the CCC are similar to those of the ICC in many circumstances as they are both estimated through variance (Carrasco and Jove, 2003; Chen and Barnhart, 2008). Perhaps then the results of the CCC would be sufficient to analyse the agreement and reliability of this data set.

6.5.3 Assessing agreement and reliability in other studies

Comparison of the agreement and reliability assessments of this data logger technique with similar choice tests from the literature is difficult. Seaman et al. (2008) conducted an experiment with a very similar set-up, but all evaluation of behaviour came from analysis of video data, and thus no parameters of between-method agreement or reliability were produced. Blom et al.'s (1992) experiment was both similarly designed, and compartment passage was remotely monitored – although with an LED sensor system in place of a data logger. However, in this paper the only comparisons between video and remote sensing methodology were Pearson's correlation coefficients. This method is not ideally suited to study agreement/reliability because it does not account for systematic bias (Hunt, 1986). Additionally, no mention of testing the normality of the data was presented.

Within the wider literature evaluating the dependability of behavioural data collected by data loggers, comparison of parameter choices favour those presented in the current study, or is often difficult due to the difference in data collection/presentation. To monitor temporal distribution in reindeer, Van Oort et al. (2004) used data loggers to monitor activity. Data logger output was compared to activity data collected through direct observation. However, as reindeer were said either to be active or inactive, the outcome variable was binary. Percentage of agreement was then used to assess agreement between the two methods. This parameter is widely criticised for its failure to discriminate between actual agreement and agreement which is due to chance (de Vet, 2005). Ledgerwood et al. (2010) used data loggers to monitor lying behaviour in dairy cattle. Data from the data logger were compared to video analysis of the same period. The resulting data were also binary, as cows were said either to be standing or lying. The results of the data logger were then compared to those of the video (which was considered the gold-standard) for sensitivity, specificity, and predictability. These are not direct measures of agreement or reliability, but do have value assessing the data set. Overall R^2 values were also presented for total lying time, and number of lying events as a measure of agreement. No mention was made of what kind of correlation coefficient was employed – presumably Pearson's. There are issues inherent in relying solely on this parameter to assess agreement, as Pearson's correlations can produce a perfect linear relationship without any actual agreement due to systematic bias (for a more detailed explanation, see Ch 2, this text).

Broadening the scope even wider to simply look for literature examining the dependability of behavioural tools/tests, it is possible to identify studies using similar approaches to measuring agreement and reliability. Temple et al. (2013) assessed the 'test-retest reliability' of the Welfare Quality® protocol. Although some terminology differed from that presented in the current study, essentially they used Bland and Altman's (1986) limits of agreement and Spearman's correlation coefficient to assess agreement, and ICC to assess reliability. While some of the specific tests used vary, they are similar in nature and follow the definitions and suggestions supplied by de Vet (2005) for assessing agreement and reliability.

6.5.4 Conclusions

The above analyses provided sufficient evidence to conclude that the use of data loggers and the data cleaning method was an adequate method to monitor entrance of the cats into and out of the plus-shaped choice chambers.

6.6 References

Bland, M.J., Altman, D.G., 1986. Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet 327, 307-310.

Blom, H.J., Van Vorstenbosch, C.J., Baumans, V., Hoogervorst, M.J., 1992. Description and validation of a preference test system to evaluate housing conditions for laboratory mice. Appl. Anim. Behav. Sci. 35, 67-82.

Blom, H.J.M., Tintelen, G., Vorstenbosch, C.J.A.H.V., Baumans, V., Beynen, A.C., 1996. Preferences of mice and rats for types of bedding material. Lab. Anim. 30, 234-244.

Carrasco, J.L., Jove, L., 2003. Estimating the Generalized Concordance Correlation Coefficient through Variance Components. Biometrics 59, 849-858.

Chen, C., Barnhart, H.X., 2008. Comparison of ICC and CCC for assessing agreement for data without and with replications. Comput. Stat. Data Anal. 53, 554-564.

de Vet, H., 2005. Observer Reliability and Agreement. Encyclopedia of Biostatistics. 1-5 .

Hunt, R.J., 1986. Percent Agreement, Pearson's Correlation, and Kappa as Measures of Inter-examiner Reliability. J. Dent. Res. 65, 128-130.

Ledgerwood, D.N., Winckler, C., Tucker,C.B., 2010. Evaluation of data loggers, sampling intervals, and editing techniques for measuring the lying behavior of dairy cattle. J. Dairy Sci. 93, 5129-5139.

Lessells, C.M., Boag, P.T., 1987. Unrepeatable Repeatabilities: A Common Mistake. Auk 104, 116-121.

Lin, L.I., 1989. A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometrics , 255.

Martin, P., Bateson, P., 2007. Measuring Behaviour : An Introductory Guide, 3rd ed. Cambridge University Press, Cambridge.

Seaman, S.C., Waran, N.K., Mason, G., Richard B. D'Eath, 2008. Animal economics: assessing the motivation of female laboratory rabbits to reach a platform, social contact and food. Anim. Behav. 75, 31-42.

Temple, D., Manteca, X., Dalmau, A., Velarde, A., 2013. Assessment of test–retest reliability of animal-based measures on growing pig farms. Livestock Science. 151, 35-45.

Van Oort, B.E.H., Tyler, N.J.C., Storeheier, P.V., Stokkan, K., 2004. The performance and validation of a data logger for long-term determination of activity in free-ranging reindeer, Rangifer tarandus L. Appl. Anim. Behav. Sci. 89, 299-308.

7 THE EFFECT OF DIFFERENT TYPES OF ENVIRONMENTAL ENRICHMENT ON THE BEHAVIOUR AND PHYSIOLOGY OF BOLD AND SHY CATS

7.1 Abstract

Environmental enrichment (EE) has been suggested as a mechanism to reduce stress in shelter cats. This study investigates the behavioural and physiological responses associated with its provision. Seventy-two cats housed singly in standard cages were allocated to one of three EE treatment groups, and provided a hiding box (BOX), a perching shelf (SHELF), or no additional EE (CTRL). Bold and shy cats were approximately balanced between treatment groups. Continuous focal observations of the activity, location in the cage, and posture were conducted from video recordings for two 4-h periods/day/cat. Food intake and Cat-Stress-Scores (CSS) were recorded daily. Faecal samples were collected for analysis of faecal glucocorticoid metabolites (FGM). These outcome variables were then analysed for the influence of treatment group, whether the cat was bold or shy, day in study, and all interactions between terms. Cats in BOX had significantly lower FGM, and consumed significantly more food daily, than did cats in CTRL. Shy cats had a significantly greater probability of registering a CSS≥3 than did bold, had a significantly greater CSS on days 1-3 than did bold cats, and within the treatment group BOX, shy cats spent a significantly greater percentage of time in the hiding box than did bold cats. Day in study was a significant factor for daily food intake and percentage of time spent eating – which tended to increase across time – and for percentage of time spent grooming – which tended to decrease across time. These results indicate caging is a stressor that can be dampened by the inclusion of EE (particularly a hiding box), that bold and shy cats appear to experience different severities of stress, and that the stress diminishes with time. Provision of a hiding box may be a simple and important strategy for helping cats cope with the stress implicit in singly housed confinement.

7.2 Introduction

When cats are placed in shelters, they often experience an initial peak in which reduces over time. Evidence for this response has been provided by studies using the Cat-Stress-Score (CSS) (Dybdall et al., 2007; Kessler and Turner, 1997; Tanaka et al., 2012), and by CSS alongside FGM and the percentages of time spent eating and grooming (Ch 3, this text), suggesting that this initial stress response stabilises after 4 days to 1 week. Other variables can also be monitored for signs of stress in shelter cats, such as changes over time in body weight and food intake (Tanaka et al., 2012).

Design of the captive environment has potential to influence the way the animal experiences its environment. The addition of certain types of environmental enrichment (EE) could potentially improve the affective state of cats, by reducing their fear, increasing their sense of control, or increasing their perceived safety. This could diminish their reaction to the introduction to a new environment, thereby reducing their initial peak in stress response. However, the type of EE must be appropriate for cats, and potentially, certain cat-level characteristics (such as age or sex) could influence the responses to enrichment. To identify what types of EE commonly employed for use with shelter cats were most likely to contribute to improving affective state, Ellis (Ch 5, this text) measured which type of EE cats allocated most of their time in proximity to in captive settings using a choice test. The assumption was that cats would interact more with EE items that would minimise a negative affective state, and/or contribute to a positive affective state (the concepts of negative and positive affective states are summarised in Fraser, 2008). Cats were found to significantly use the compartment containing a hiding box over both the empty compartment and the compartment containing a toy. The compartment containing an elevated shelf was the second most used compartment. There are some problems inherent in this type of test. For example, if cats are most likely to sleep in the hiding box compartment and play in the toy compartment, then perhaps this usage is simply reflections of the cat's time budget. However, investigation into the effect of providing different types of EE on indicators of stress can demonstrate the importance of this time allocation to the welfare of the cat.

There is also the potential that the way shelter conditions are experienced depends on one or more cat-level characteristics. Dybdall et al. (2007) found that cats surrendered by their owner registered significantly higher CSS than did cats entering the shelter as strays, presumably because they perceived the conditions as a greater danger, and this affected their affective state. The authors of that article suggest that cat temperament – or behavioural style – could also have a great influence on stress as a result of shelter conditions. Of the behavioural styles commonly studied in cats, shyness/boldness could be the key behavioural style to influence a cat's response to the stressors implicit in being housed in a shelter. McCune (1995) defines the bold/shy behavioural style as a general response to novelty irrespective of whether the novelty is human or object. As one of the primary stressors of shelter environments is exposure to the unfamiliar, according to this definition it follows that bold cats and shy cats would have greatly different experiences in shelter settings, and could potentially benefit from different management and husbandry conditions. Additionally, while Ellis (Ch 5, the text) found no major differences in EE usage between cat-level characteristics, there was some evidence that shy cats may have used the hiding box more than bold cats. It is therefore possible that cat-level characteristics – such as behavioural style, particularly bold vs. shy – might influence the perception of shelter conditions, the affective states experienced, the stress expressed, the most beneficial type of EE, and ultimately, the welfare of the cat.

The goal of this chapter was to investigate stress responses in singly housed domestic cats provided with different EE treatments. Based on the previous research in this thesis, it was hypothesized that 1) cats in the treatment group containing the EE type cats previously allocated the most time in proximity to (BOX) would have lower indicators of stress than those in either of the other treatment groups (SHELF and CTRL), 2) that the magnitude of the stress response would vary based on whether a cat is bold or shy, and 3) across treatment groups, each indicator of stress would reduce over time, and stabilise after a period of approximately 4 days to 1 week.

7.3 Methods

7.3.1 Subjects

This study was conducted using 72 domestic cats obtained from a local animal shelter, of which 35 were males (26 were intact and 9 were neutered) and 37 were females of unknown reproductive status. Forty-five animals came into the shelter as strays and 27 were surrendered to the shelter by their owner. All of the cats were judged to be adults by the absence of kitten teeth. Thirty-three of the cats were judged to be 2 years of age or younger, while 39 of the cats were judged to be over 2 years of age. Using the emergence test with a cut-point of 10 s at the Humane Society (as described in chapter 4), 41 cats were classified as bold, while 31 were classified as shy. The mean length of time at the shelter before transport approximately 5 km to the study site was 5.2 ± 2.46 (±SD) days. Before inclusion in the study, all cats were deemed by shelter staff to be healthy, and were given a veterinary examination upon arrival at the university, and again before return to the shelter after the study's completion. All of the cats used in this study were also used concurrently in the study described in the subsequent chapter. At any one time, there were between 3 and 6 cats in the study.

7.3.2 Housing and management

All cats were individually housed in stainless steel cages (58 cm x 79 cm x 79 cm) for 10 days. Within each cage, there was a litter pan (33 cm x 28 cm x 13 cm) and food and water dishes secured to the cage door. Cats were offered 100 g of dry cat food (50 g Adult Indoor, Nutrience by Hagan; and 50 g Friskies by Purina) at 09:00 h daily. Daily weights and food intake were recorded, as well as the presence or absence of urine, faeces, and/or vomit. Occasionally, if food intake had been exceptionally low for several days, the diet of individual cats was supplemented with wet food. Litter pans were cleaned and water bowls were filled by animal care technicians twice daily: once between 09:00 h and 10:00 h, and once between 15:00 h and 16:00 h. The room was maintained at a temperature of 20^{o}C (range ± 2 ^{o}C). The researcher and animal care staff interacted with the cats only as much as was necessary in the routine feeding, cleaning, and experimental protocol.

The cages were in two banks of three cages that faced each other so that all cats could see, hear, and smell each other. Cats were provided with 12 h of fluorescent lighting and 12 h of darkness with infrared illuminated lighting. The fluorescent lights came on at 06:00 h and went off at 18:00 h each day. Both banks of cages had three black-and-white CCTV cameras (Panasonic, Germany) and two infrared illuminated microlights (880 Infra Red Illuminator, Dennard, UK) on top of them focussed on the opposite cages so that each cage had one dedicated camera, and the light was positioned so that each cage was sufficiently illuminated to permit behavioural observation.

7.3.3 Experimental treatments

Cats were randomly allocated to one of three treatment groups containing different environmental enrichment items. Treatment group 1 (BOX) provided cats with a hiding opportunity in the form of a cardboard box (26 cm x 26 cm x 66 cm) with a semicircular entrance hole (20 cm x 16 cm) cut in the front. Cats could go inside this box, but not on top. Treatment group 2 (SHELF) provided cats with a perching opportunity in the form of a cardboard box shelf (55 cm x 20 cm, 32 cm off the floor). Treatment group 3 (CTRL) was a control group that provided cats with no additional environmental enrichment items. The cats were free to shift any movable objects throughout the day, but twice a day when cages were cleaned the items were rearranged. There were 24 cats in each treatment group; Table 7.1 presents the distribution of cat-level characteristics between them.

Table 7.1 Distribution of cats among the treatment groups (n=72)

	No. of cats								
	Sex		Stray/ Surr		Age		Behavioural style		LOS (days)
	M	F	Stray	Surr	≤2 y	>2 y	Bold	Shy	mean (SD)
BOX	12	12	17	7	12	12	15	9	5.4 (2.10)
SHELF	8	16	14	10	12	12	12	12	4.6 (1.74)
CTRL	15	9	14	10	9	15	14	10	5.5 (3.28)

M=Male
F=Female
Surr=Surrendered
y=years
LOS=Length of stay at animal shelter before study

7.3.4 Continuous quantitative behavioural observations

Cats were time-lapse recorded (VCRs: CTR-3024, Computar, UK; multiplexer: Sprite dx, Dedicated Micros, UK) 24 h a day for 9 days. Two 4 h periods (morning period: 07:00-11:00 h, evening period: 23:00-03:00 h) of each of the nine days were analysed for each cat's general activity, location within the cage, and posture using continuous observation (Martin and

Bateson, 2007). These time windows were chosen to optimise observation of eating (morning period) and grooming (evening period), which were identified as behaviours that changed in association with habituation (Ch 3, this text). Recordings were made at 50 frames/s, and viewed at 10 times the actual speed. The ethogram for location, posture, and activity is given in Table 7.2. A behavioural bout was considered to have concluded if the animal ceased to perform the behaviour for at least 5 s. The analysis of behaviour was conducted by two different observers, using Noldus Observer 5 (Noldus Information Technology, Wageningen, The Netherlands). Of the total 4928 h of recordings watched, one observer analysed 3828 h of footage, while the other analysed the remaining 1100 h. Observations were distributed so that no cat ID, day in study, or period (morning or evening) was conducted by only one observer.

7.3.5 Food intake

Before daily feedings, the dry food remaining in the dish from the previous day's feeding was weighed. Consumption of wet food was not measured. The weight of each cat was recorded on day 1 and day 10.

7.3.6 CSS recordings

Once per day (at 11:00 h) on days 2-9, the researcher stood in front of the cats cages for 10 min in order to habituate the cats to the observer presence, and then assigned the cats a score from 1 (fully relaxed) to 7 (terrorised) in 11 behavioural/postural categories (Kessler and Turner, 1997). The assessment for each cat was then repeated within a 15 min interval.

Table 7.2 Ethogram for quantitative behavioural observations (adapted from UK Cat Behaviour Working Group, 1995)

Behavioural Category	Behaviour	Definition
Activity	Resting	Cat remained generally inactive
	Grooming	Cat licked its body or licks its paw and passed the paw over its head
	Eating	Cat consumed (or appeared to consume) food
	Drinking	Cat lapped water
	Manipulating	Cat manoeuvred or attempted to manoeuvre an object with its paw
	Locomotion	Cat moved position within its enclosure
	Out of cage	Cat was not present in cage
	Other activity	Any activity not defined above
Location	On shelf	Cat was positioned on top of the provided shelf
	In litter pan	Cat was positioned in the provided litter pan
	Behind litter pan	Cat was positioned behind the provided litter pan
	Cage floor	Cat was positioned on the cage floor
	Out of cage	Cat was not present in cage
	Other location	Any location not defined above
Posture	Lying	One side of cat was in complete contact with the ground
	Sitting	Pads of the front paws were on the ground with the front legs straight and the rump on the ground
	Standing	Cat was positioned with four paws on the ground, rump raised
	Locomotion	Cat moved position within its enclosure
	Out of cage	Cat was not present in cage
	Other posture	Any posture not defined above
View obscured	View obscured	Sight lines to the cat were obstructed

7.3.7 FGM measurement

Faecal collection

When litter pans were cleaned all faecal samples present were homogenised and placed into 30ml Nalgene tubes. The samples were stored at -4^{o}C for up to 21 h and then stored at -20^{o}C until extraction. An average of 6.4 ± 2.75 samples were collected from each cat.

Hormone extraction and analysis

Glucocorticoid metabolites were extracted and analysed from the faecal samples according to Möstl and Palme's (2008) protocol using EIA 9.3 11-oxoaetiocholanolone (Lab-code: 72-alt; EIA first described: Palme and Möstl (1997), and previously validated for use in cats (Palme et al., 2001). In the current study, inter- and intra-assay coefficients of variation were 14.4% (n=24) and 4.3% (n=975) respectively.

7.3.8 Statistical analyses

Inter-observer agreement

The inter-observer agreement of the two observers was evaluated for both durations and frequencies per sequence (see Ch 2, this text) of behaviours from 30 periods (4 h), from six different cats, distributed across nine days and morning/evening sample periods, using methods of percentage of agreement, Pearson's *r*, and Cohen's kappa test statistic (Jansen et al., 2003). Since these are the three most common agreement parameters presented throughout the literature, all have been included for cross-study comparison. Minimum acceptability of results was set at ≥80% for percentage agreement (Cooper et al., 2007), ≥0.8 (or a strong effect) for Pearson's r (Ferguson, 2009), and >0.6 (substantial agreement or better) for Cohen's kappa test statistic (Landis and Koch, 1977).

Analysis of behavioural variables

For each behavioural class, percentage of time spent performing each of the specific behaviours in each 4 h period were calculated, and daily median and IQR values were produced for each cat. As all locations were not present in all groups, separate summary statistics for location were calculated for each EE treatment group.

The data for each behaviour were analysed by a linear mixed model with the fixed effects of day in study, treatment group, bold/shy behavioural style, and all possible interaction combinations of the preceding variables, with an arma (1,1) within-cat correlation structure to account for repeated measures on cats (Dohoo et al., 2009). Data were transformed where appropriate (details provided in Table 7.7) to meet the model assumptions of normality and homoscedasticity of the residuals. This analysis was performed only with variables that had a median value of <99% and greater than 1%, as transforming data to satisfy model assumptions was impossible in variables with extreme distributions. A model was then tailored for each specific behaviour by removing non-significant factors until all remaining terms were either significant, part of a significant interaction term, or were one of the key variables related to the research question (i.e. day in study, treatment group, or bold/shy behavioural style). For factors found to be significant, Bonferroni-adjusted pairwise comparisons were used to compare categories. Where the Bonferroni pairwise adjustments for a large number of tests resulted in non-significant results, the results from unadjusted comparisons were presented.

Analysis of non-behavioural variables

All samples collected during the twice daily litter pan cleanings were analysed for FGM concentrations. If more than one faecal sample was produced by a cat in one day, an average of the FGM concentrations was calculated (Touma and Palme, 2005). Within each cat, FGM concentrations from each sample were averaged, and a median and IQR across all cats was calculated for overall summary statistics. For each cat, median daily amounts of dry food eaten and body weight change from day 1 to day 10 were also calculated. A mixed model was then built (in the same way as described above for behavioural results) for daily food intake, a similar models were built for FGM and body weight change, but the model for FGM adjusted for food intake the day before, and the model for body weight change did not include day in study as a factor. Transformations were used where appropriate to meet model assumptions.

Analysis of CSS

Both daily CSS values were averaged to produce a single daily score. These daily CSS values were averaged within cats to produce a single score for each cat, and a median and IQR for these values were calculated to produce overall descriptive statistics. Although the results from the CSS are ordinal and thus should technically not be analysed using the above model, the creators of the system advocate assuming changes within this small range approximate a linear system (Kessler and Turner, 1999) and thus this analysis was included. They also suggest that scores up to 3 reflect cats experiencing "generally acceptable" levels of stress, while scores greater than 3 reflect cats experiencing "unacceptable" levels of stress (Kessler and Turner, 1999) (although Ottway and Hawkins (2003) defined CSS <4 as "welfare unaffected" and CSS≥4 as "poor welfare" or "very poor welfare"). For this reason, daily CSS results have been analysed as both a continuous outcome variable and as dichotomous, with results classified as either <3 or ≥3. The distribution of CSS in these newly created categories between treatment groups and bold/shy categories is given in Table 7.6. Continuous CSS results were analysed using the same mixed model and post-hoc testing described above. Dichotomous CSS results were analysed using a generalised estimating equation (GEE) method (Hanley et al., 2003), and model building was carried out in a similar way to the linear model. The low number of cats registering a CSS ≥3 imposed limitations on the GEE method; therefore an additional GEE with a cut-point of ≥ 2.5 (which had a more even distribution of scores) was carried out.

7.3.9 Ethical approval

This project was approved by UPEI's Animal Care Committee under protocol number 09-051, and followed the guidelines of the Canadian Council of Animal Care's "Guide to the Care and Use of Experimental Animals".

7.4 Results

The inter-observer agreement of the two observers satisfied all of the minimum acceptability standards for duration of behaviours for all three measures of agreement set out in the methods section, while the values for frequencies failed to meet the standards for any of the three measures (Table 7.3). As a result, all subsequent analyses were only conducted on the durations of the behaviours.

Table 7.3 Mean (±SD) inter-observer agreement between 2 observers of 30 four-h observations

	Percentage agreement	Pearson's *r*	Cohen's kappa
Duration/sequence based†	89 (12)	0.99 (0.04)	0.63 (0.27)
Frequency/sequence based†	57 (13)	0.77 (0.20)	0.46 (0.13)

†See Ch 2 (this text) for an in depth discussion of sequence based inter-observer agreement analysis

The overall median and IQR of each behaviour are displayed in Figure 7.1, while the median and IQR for FGM, daily food intake, percentage of body weight change, and CSS are given in Table 7.4. Table 7.5 contains the breakdown of the dichotomised CSS variable between treatment groups and bold/shy behavioural styles. The influence of day in study, treatment group, and bold/shy behavioural style had on select behaviours (those found to have significant relationships), FGM, daily food intake, percent body weight change, and CSS can be seen in Table 7.6.

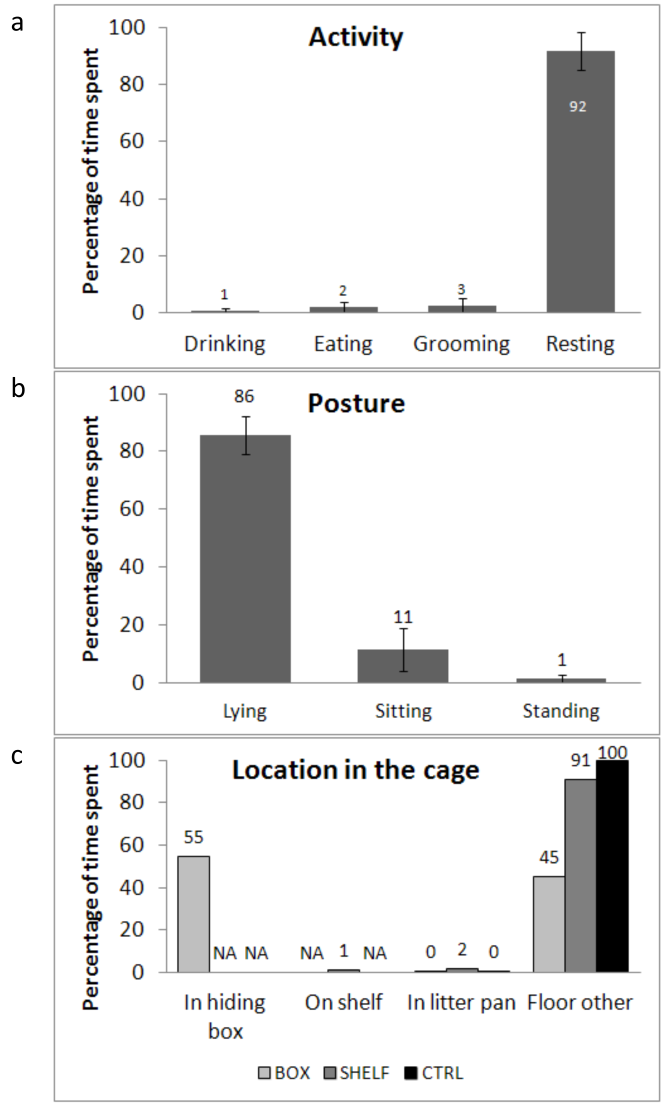

Figure 7.1 Median and interquartile range of percentage of time spent in each behaviour of the three behavioural categories (n=72): activity (a), posture (b), and location in the cage (c). Location in the cage is presented with the treatment groups separated, as availability of locations varied based on treatment groups. NA indicates where locations were not available to treatment groups. Behaviours with median values <1 were not included on graphs for any of the behavioural categories. The numbers associated with each bar indicate the percentage that bar represents.

Table 7.4 Median, 1st and 3rd quartile for faecal glucocorticoid metabolite (FGM) concentrations (ng/g), daily food intake (g), percent body weight change, and Cat-Stress-Score (CSS) (n=72)

Variable	Median (Q_1, Q_3)
FGM (ng/g)	384 (131, 881)
Daily food intake (g)	51.2 (26.4, 86.4)
Percent body weight change	0.5 (-5.3, 4.9)
CSS	2.4 (2.1, 2.5)

Table 7.5 Distribution of total Cat-Stress-Score (CSS) <3 or ≥3 between treatment group and bold/shy behavioural styles (n=72).

	Treatment group	CSS <3	CSS ≥3
Bold	BOX	125	0
	SHELF	92	4
	CTRL	108	3
Shy	BOX	38	25
	SHELF	58	35
	CTRL	48	30

Note: values represent number of times a CSS < or ≥ 3 was registered by any cat on any day, separated by treatment group and whether the cat was bold or shy.

There were significant effects of environmental enrichment on FGM (P=0.04) and daily food intake (P=0.03) (Figure 7.2). BOX was significantly different than CTRL, while SHELF was not significantly different from either of the other groups. There was a tendency (P=0.05) for a treatment effect on percentage body weight. Cats in treatment group N (mean=3.38, SD=7.70) lost more weight than cats in treatment group H (mean=-2.47, SD=6.24).

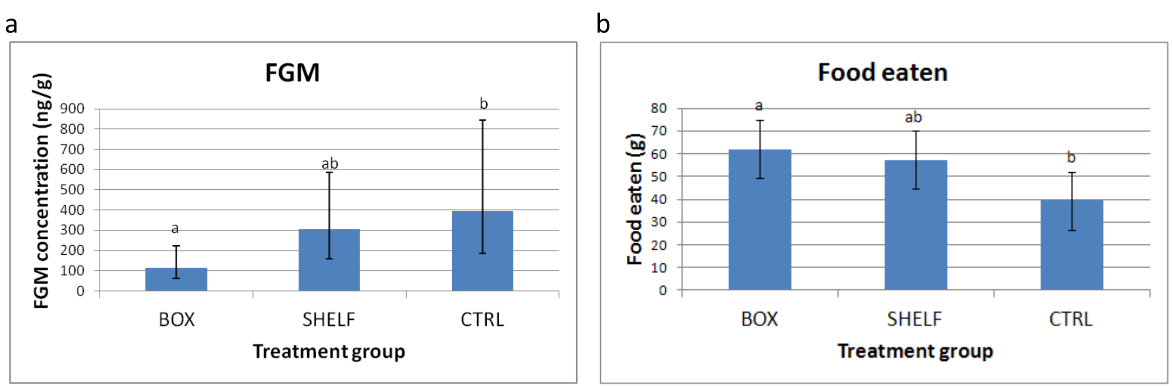

Figure 7.2 Faecal glucocorticoid metabolite (FGM) concentration (a) and amount of food eaten per day (b), separated by treatment group (n=72). Graphs are of means and error bars are confidence intervals. FGM has been back-transformed. Results presented for amount of food eaten were Bonferroni-adjusted. Treatment groups with the same letter code were not significantly different (P>0.05).

Table 7.6 Final models, transformations, and significant values for all behaviours with medians <99% and >1% of each behavioural class, faecal glucocorticoid metabolite (FGM) concentrations, daily food intake, percentage of body weight change, and Cat-Stress-Score (CSS) as both a continuous and dichotomous outcome variable (n=72)

	Lambda	EE	BS	Day	EE*BS	Day*EE	Day*BS	EE*BS*Day	FIDB
Eating	0.43	†	$F_{1,67.2}$=3.22, P=0.08	$F_{8,295}$=2.73, P=0.01	-	-	-	-	NA
Grooming	0.41	†	†	$F_{8,285}$=2.05, P=0.04	-	-	-	-	NA
Resting	9	$F_{2,64.4}$=5.07, P=0.01	†	†	†	†	$F_{8,282}$=2.59, P=0.01	$F_{16,282}$=1.88, P=0.02	NA
Standing	0.37	†	†	$F_{8,275}$=2.47, P=0.01	-	-	-	-	NA
Sitting	0.5	†	†	†	$F_{2,65.6}$=4.06, P=0.02	-	-	-	NA
Lying	5	†	†	†	†	$F_{16,284}$=1.59, P=0.07	$F_{8,284}$=2.08, P=0.04	$F_{16,284}$=1.72, P=0.04	NA
Hide box (BOX)	7	NA	$F_{1,22.6}$=5.33, P=0.03	†	NA	NA	-	NA	NA
Cage floor (BOX)	0.5	NA	$F_{1,23.2}$=4.17, P=0.05	†	NA	NA	-	NA	NA
On shelf (SHELF)	0	NA	†	†	NA	NA	-	NA	NA
Cage floor (SHELF)	9	NA	†	†	NA	NA	$F_{2,67}$=2.13, P=0.04	NA	NA
Litter pan (SHELF)	0	NA	†	†	NA	NA	-	NA	NA
FGM	0	$F_{2,52.5}$=3.50, P=0.04	†	†	-	-	-	-	$F_{223,66.5}$=1.47, P=0.03
Food intake	1	$F_{2,70.1}$=3.69, P=0.03	†	$F_{8,301}$=3.79, P<0.01	-	-	-	-	NA
% body weight change (day 1-10)	1	$F_{2,68}$=4.07, P=0.02	†	NA	-	-	-	-	NA
CSS continuous	0.5	†	$F_{1,68.7}$=12.66, P<0.01	$F_{7,341}$=2.82, P<0.01	-	-	$F_{7,341}$=3.52, P<0.01	-	NA
CSS dichotomous (≥3)	binary	†	χ^2_1=28.39, P<0.01	χ^2_7=21.48, P<0.01	NP	χ^2_{14}=31.03, P<0.01	NP	NP	NA

Lambda=power transformation
EE=treatment group
BS=bold or shy cat
FIDB=Food intake the day before
-=removed during model building process because p≥0.05
NA= not applicable to the particular model
NP=not possible: due to the low number of bold cats with CSS≥3, interaction with BS could not be explored
†=included in the model, but P-value >0.1

For cats in BOX, the bold/shy behavioural style was found to be a significant factor for percentage of time spent in the hiding box, and for the probability of registering ≥3 on the CSS. Shy cats were significantly more likely to spend a higher percentage of time in the hiding box, and had a significantly higher probability of registering ≥3 on the CSS (Figure 7.3).

Day in the study was found to be a significant factor for percentage of time eating, grooming, and standing, and the amount of food eaten. Daily estimates, confidence intervals, and significant differences for percentage of time spent eating, percentage of time spent grooming, and food eaten (g) are shown in Figure 7.4; the results for percentage of time spent standing can be found in Appendix B.1. Both the percentage of time spent eating, and the daily amount of food eaten, increased over time. The percentage of time spent grooming reduced after an initial spike. The percentage of time spent standing was stable except for a spike on day three.

a b

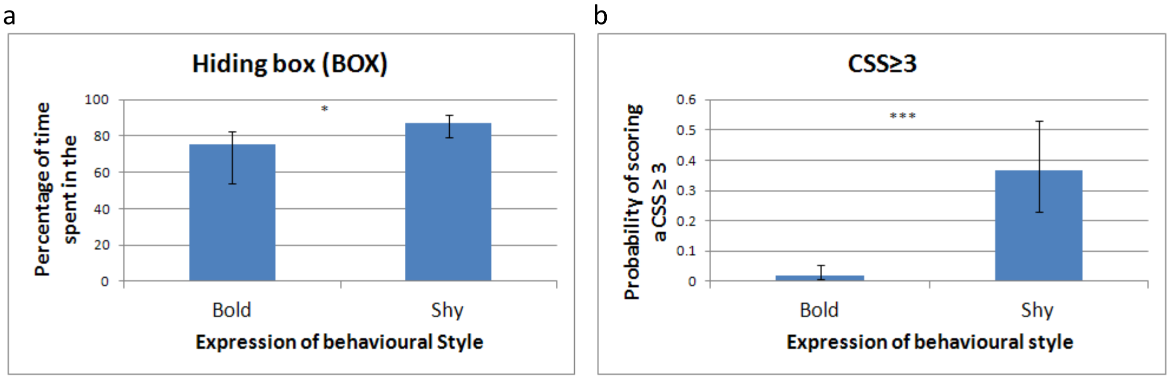

Figure 7.3 For cats in BOX, percentage of time spent in the hiding box (a) and probability of registering a Cat-Stress-Score (CSS) ≥3 (b), separated by bold/shy behavioural style (n=72). Graphs are of means and error bars are confidence intervals. Percentage of time spent in the hiding box has been back-transformed. Astericks indicate significant difference between bold and shy cats. One asterisk means P<0.05, and three asterisks mean P<0.001.

The interaction between treatment group and the bold/shy behavioural style was a significant factor for the percentage of time spent sitting. There was no significant difference between treatment groups in this variable for bold cats, but shy cats in BOX spent significantly more time sitting than those in SHELF (P<0.05) or CTRL (P<0.01). For BOX only, bold and shy cats differed in the percentage of time spent sitting (P<0.01). These results are graphically represented in Appendix B.2.

a b c

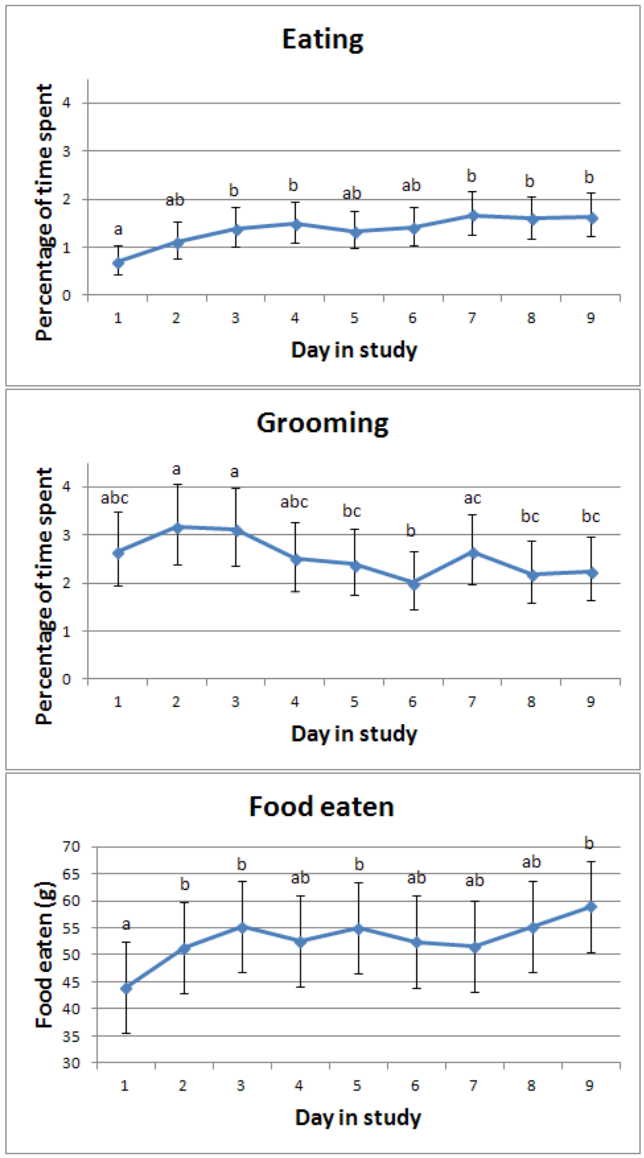

Figure 7.4 Effect of day in study on percentage of time spent (a) eating and (b) grooming, and (c), weight of food eaten (n=72). Graphs are of means, and error bars are confidence intervals. Percentages of time spent eating and grooming has been back-transformed. Results presented for percentage of time spent eating and amount of food eaten daily were Bonferroni adjusted, results for percentage of time spent grooming were not Bonferroni-adjusted. Treatment groups with the same letter code were not significantly different (P>0.05).

The interaction between day and the treatment group was a significant factor for CSS as a dichotomous variable. While this term makes no mention of behavioural style, Table 7.5 and Figure 7.3b show that there are very few scores ≥3 registered by bold cats. Therefore detailed results are presented in Figure 7.5a-c for shy cats only, with shy and bold cats shown together in Figure 7.5d. The probability of registering a CSS≥3 does not change across time in BOX, but reduces with time in both SHELF and CTRL.

Unfortunately, for CSS as a dichotomous variable, it was not possible to consider interactions with the bold/shy behavioural style because in some of the treatment groups there were no bold cats that registered CSS ≥3 (Table 7.5). A regular logistic regression (not clustering for cats) found this interaction to be significant, but it is likely that the effect of the clustering would be quite strong. Pairwise comparisons presented in Figure 7.5 were conducted for bold and shy cats separately, in order to consider the interaction between treatment group and day in study. An additional GEE was carried out at a cut-point of CSS ≥2.5 – which had a more even distribution, forcing more bold cats to register scores above the cut-off point – and none of the two-way interactions were significant. The three-way interaction between day, treatment group, and bold/shy behavioural style was significant, but upon closer examination of the relationships, this effect seemed to be random variations in the data as no obvious patterns emerged. This was taken as evidence that it was acceptable to look at dichotomous CSS results without considering the interactions with the bold/shy behavioural style.

The interaction between day and the bold/shy behavioural style was a significant factor for CSS as a continuous variable, and the percentage of time spent on the floor of the cage (for cats in SHELF). Continuous CSS was stable across time in bold cats and exhibited a clear reduction across time in shy cats. It was significantly different between bold and shy cats on days 2, 3 and

4 (Figure 7.6). Within cats in SHELF, bold cats seemed to spend longer on the floor after day 1, while the pattern of floor use in shy cats was stable or slightly decreasing with time, except for day 3. However, the pattern was not clear in either type of cat. Bold and shy cats in SHELF spent a significantly different amount of time on the floor on day 3. These results are graphically represented in Appendix B.3.

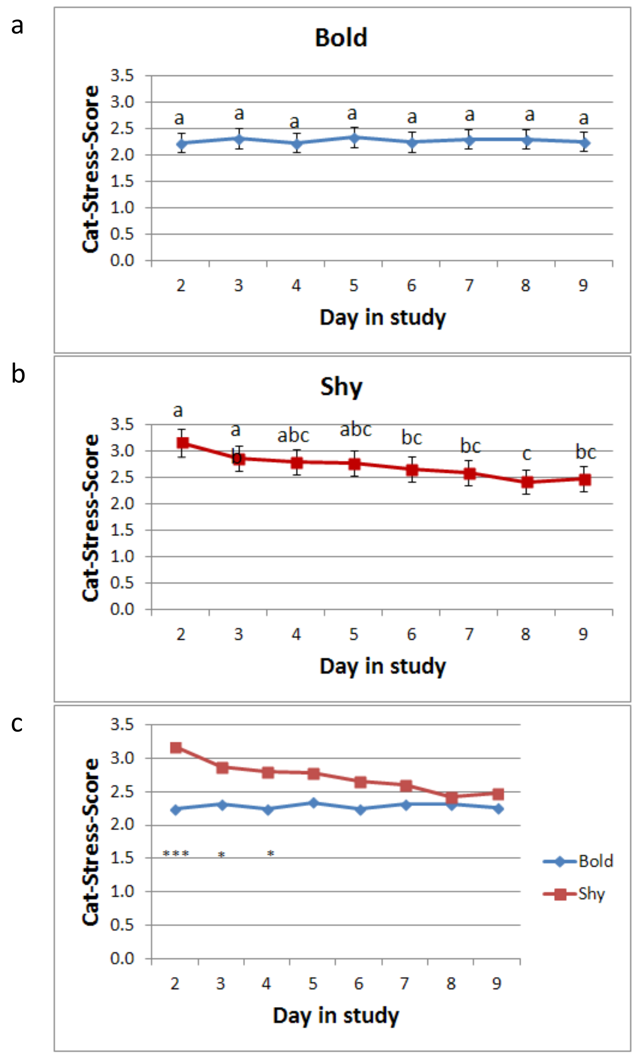

Figure 7.6 Effect of day in study on Cat-Stress-Score (CSS) as a continuous variable for bold and shy cats (n=72). Results presented are of back-transformed means, and the error bars are 95% confidence intervals. Figure 7.6a presents the changes in bold cats over time, and Figure 7.6b shows the changes in shy cats over time. In Figures 7.6a and 7.6b treatment groups with the same letter code were not significantly different (P>0.05). Figure 7.6c shows the differences between bold and shy cats over time. This figure uses asterisks to indicate significant differences between bold and shy cats on particular days. One asterisk means P<0.05, and three asterisks mean P<0.001. Results presented were Bonferroni-adjusted.

The three-way interaction between day in the study, treatment group, and the bold/shy behavioural style was a significant factor for the percentages of time spent resting and lying. For bold cats, the percentage of time spent resting declined in all treatment groups on days 1-3, it failed to rebound only for cats in BOX, which became significantly lower than both other treatment groups by day 8 ($P<0.05$). Any patterns in the differences between treatment groups in shy cats were less clear and were interpreted as random noise. There were no significant differences between bold and shy cats on any of the days in any of the treatment groups. These results are graphically represented in Appendix B.4. Although the percentage of time spent lying produced some significant differences over time with, in particular, treatment group and bold/shy subgroups, none of the results revealed clear patterns, and were interpreted as random noise. These results are graphically represented in Appendix B.5.

7.5 Discussion

Treatment group, bold/shy behavioural style, and day in the study each influenced different indicators of stress in singly housed cats, indicating that overall, caging is a stressor that can be dampened by the inclusion of EE, that this stress is experienced differentially by bold and shy cats, and that it is a stressor that diminishes with time.

The first hypothesis of this chapter was that cats in the treatment group containing the EE type cats previously allocated the most time in proximity to (BOX) would have lower indicators of stress than those in either of the other treatment groups. For the variables FGM concentration and amount of food eaten per day, BOX was found to be significantly different from CTRL, but SHELF was not significantly different from either group. The provision of a box – the EE type used most in Ch 5 (this text) – significantly reduced the stress response of the cats as compared with a relatively un-enriched CTRL compartment, but not necessarily the other enriched compartment, SHELF. Treatment group was also a significant factor for CSS as a dichotomous variable, but only in combination with ‘day in study’. Although there was no significant difference between the treatment groups outside of the interaction term, each treatment group acts very differently across time. In the BOX treatment group, there were no significant differences across time, suggesting that no reduction in stress occurred in association with habituation– implying that cats in this treatment group experienced low stress levels, similar to the eventual baseline, right from the start. Both other treatment groups declined from day 1 until there were no further significant differences after day 6 (SHELF) and after day 4 (CTRL), suggesting that this is the time that the baseline level was reached, and thus habituation was complete. As BOX cats had a low probability of registering a CSS≥3 immediately after introduction which did not change over time, and both the other treatment groups started with higher probabilities of registering a CSS≥3 which did reduce over time, this was interpreted as evidence that providing hiding opportunities to singly housed cats can help lower stress. The percentage of time spent resting had a significant three-way interaction term between treatment group, behavioural style, and day in study. Resting behaviour reduced over time in BOX, becoming significantly lower than both other treatment groups by day 8. For shy cats, this behaviour in the BOX treatment was significantly lower than at least one of the other treatment groups in 4/9 days. Reduction of excessive amounts of time resting can be interpreted as a sign of positive psychological well-being, through a reduction of boredom (Wemelsfelder, 2005). Given this body of evidence, it is reasonable to suppose that the BOX treatment group is more likely to contribute to lower stress in cats housed in a shelter setting.

The association between the provision of a hiding opportunity and a reduction in stress is in agreement with the results of previous studies. Kry and Casey (2007) found a significant reduction in CSS between all study days for cats provided a BC SPCA Hide & Perch Box™, while the pattern for control cats was a less consistent reduction with periodic increases. They also found that cats provided the EE were significantly more likely to approach a person in an approach test, and displayed relaxed behaviours much more frequently. Godijn (2013) found that cats provided BC SPCA Hide & Perch Box™ experienced a reduction in CSS quicker than cats in the control group. The groups had significantly different CSS on days 3 and 4, but this difference disappeared on the final two test days when control cats seemingly reached the CSS baseline and were not significantly different from enriched cats. Godijn also found that cats not provided with a hiding box spent significantly more time attempting to hide behind their litter box. Carlstead et al. (1993) and Kry and Casey (2007) also found that cats not provided a hiding opportunity were significantly more likely to attempt to hide behind their litter box. Additionally, Carlstead et al. found that hiding behaviour was negatively correlated with the cortisol to creatinine ratio, suggesting that hiding may be an important mechanism for coping with the stress of captive conditions.

The second hypothesis of this chapter was that the magnitude of the stress response would vary based on whether a cat is bold or shy. Whether a cat was identified as bold or shy was a significant factor for percentage of time spent in the hiding box (for cats in BOX) and for CSS as a dichotomous variable. Shy cats spent a greater percentage of time in the hiding box, and had a much greater probability of registering a CSS$\geq$3 than did bold cats. The division of bold/shy rested on latency to emerge from a cat carrier, and obvious parallels can be drawn between this and the percentage of time spent in a cat carrier, if one makes the assumption that in both instances, the cat occupies the hiding box/carrier in an attempt to shield itself from perceived danger. Furthermore, the agreement between the CSS and the bold/shy divisions may be a result of the design of the CSS, which Fraser (2008) deems "designed more specifically to assess the degree to which cats have adapted to an unfamiliar environment, rather than the much larger class of challenges (cold, heat, injury, disease) that are commonly assumed under stress". As the bold/shy behavioural style was defined in Ch 4 and Ch 8 (this text) as the general response to novelty, irrespective of whether the novelty is human or object (McCune, 1995), there are clear parallels with what the CSS was designed to measure. In fact, the CSS itself may prove an efficient way to assess the bold/shy behavioural style. While the findings of neither of these variables are surprising, it lends some validity to the methods that they are in agreement. Whether a cat was bold or shy was also a significant factor for CSS as a continuous outcome – although as part of an interaction term with day in study. Bold and shy cats were significantly different only for days 2-4, after which the shy cats joined the bold cats in exhibiting stable CSS. This was interpreted as evidence that bold and shy cats experience the stress of shelter settings differently, and that shy cats feel the effects more strongly – especially initially – and that perhaps shy cats would benefit most from a tailored EE program.

The final hypothesis for this chapter was that across treatment groups, each indicator of stress would reduce over time and stabilise after a period of approximately 4 days to 1 week. Several of the variables examined did change significantly over time. Percentage of time eating increased with time, and percentage of time grooming decreased, which is similar to the findings in Ch 3 (this text). In addition, the amount of food eaten daily significantly increased with time. Unfortunately, determining whether these variables stabilised after 4-7 days was difficult. Neither percentage of time spent eating, nor weight of food eaten, showed significant changes from day 2-9, but some of those days failed to be significantly different than day 1 as well. Similarly, for grooming, there were no significant differences among days 4-9, but some of those days failed to be significantly different from earlier days. Furthermore, it is not possible to see if these trends continue further in time. Perhaps if data collection had continued for a greater length of time further elucidation of the patterns of these variables may have been possible. The relationship between food consumption and time may help to explain why the median percentage of body weight change was close to zero, as cats ate little at first, and then gradually started to eat more. However, there was large individual variability in this figure. Weight loss has been cited as a problem in shelter cats and has even been suggested as “a practical, indirect measure of stress and general health in shelter cats” (Tanaka et al., 2012), and while this may prove to be a valuable monitoring tool, we have no evidence to support this from the current data. Day in study was also a significant factor for CSS as a dichotomous outcome in the interaction between ‘day in study’ and ‘treatment group’, and for CSS as a continuous variable in the interaction between ‘day in study’ and ‘mode of behavioural style’. As discussed above, CSS as a dichotomous outcome reduced significantly across time in two of the three treatment groups. In these treatments, there were no further significant differences in the probability of registering a CSS≥3 after day 6 (SHELF) and day 4 (CTRL), suggesting that this is the time the baseline level was reached, and thus habituation was complete. A similar picture is shown when looking at the interaction between ‘day in study’ and ‘mode of behavioural style’ for CSS as a continuous variable. CSS as a continuous variable reduced across time for shy cats; days 6-9 were all significantly lower than day 1, and were not significantly different from each other. Given this body of evidence, it is reasonable to suppose that in a shelter setting, stress experienced by cats reduces over time with habituation, and stabilises after a period of 4 days to 1 week.

Although CSS was analysed as a continuous variable (in addition to a dichotomous variable) because the creators of the system advocate assuming changes within this small range approximate a continuous system (Kessler and Turner, 1999), there is some debate in the literature about the appropriateness of this approach. Casey and Bradshaw (2005) contend “it is unclear to what extent the intervals between the seven levels are equivalent… At present, it may therefore be more prudent to use non-parametric statistics… when comparing CSS values that vary widely”. While it is possible that the continuous analysis of the CSS in the current study was not an appropriate approach, it is notable that the results were very similar to the dichotomous analysis, and the CSS values in the current study did not vary widely. These arguments suggest that, at least for the current data set, perhaps this small range does approximate a continuous system.

There were a few behaviours that did not have a significant contribution from any of the factors investigated, notably percentage of time spent on the shelf (SHELF). As whether a cat was bold or shy had such a strong relationship with EE use in BOX, the assumption could be made that behavioural style is linked to EE use in general. However, as the use of the shelf was not different between bold and shy cats, it seems more likely that the EE in BOX (the hiding box) was relevant to the factors that differentiate between bold and shy, and perhaps the two types of cats were using the hiding box in different ways, serving different psychological functions (i.e. a shy cat uses it to hide from perceived danger, while a bold cat benefits only from the reduced boredom of additional enclosure complexity). Even though the shy cats used the hiding box significantly more than the bold cats, the usage by either type of cat was much greater for the EE in BOX than it was for the EE in SHELF, and although FGM concentrations and food intake were not significantly different between BOX and SHELF, BOX was significantly greater than

CTRL, while SHELF was not. As there was only minimal use of SHELF and a lack of significant difference in the potential indicators of stress from an empty control compartment, there was no evidence that the addition of perching opportunities was beneficial in reducing stress for a singly housed cat. However, it is possible that there was some problem with the specific shelves used in this study, such as size or material. Smith et al. (1994) showed that in an outdoor platform structure, cats preferred a wooden surface over metal, plastic, or carpet. They hypothesised that the carpet (a highly used substrate indoors) was used less frequently outdoors because of water logging due to weather. It is possible that there was something aversive about the shelf offered in the SHELF treatment, as the median time spent on it was 1.14%, and the median time spent on the factory standard shelf used in Ch 3 was 51%. It is possible that this large discrepancy is a result of different sample periods or numbers of subjects, but the magnitude of the difference suggests something else was at work, especially in light of the scientific and anecdotal reports of high levels of use of elevated platforms by caged cats (Deluca and Kranda, 1992; Podberscek et al., 1991; Smith et al., 1994). Although part of the purpose of this work was to tease out differences in the effect of different types of EE, the discrepancies between the use of the shelf here and in Ch 3 suggest that more work is needed to fully understand the potential effects of the shelf, especially if size and/or material of the shelf are critical to the nature of this relationship (e.g. if the shelf size in the current study was too small for cats, or if another surface material would have been more appealing).

As was the case in Ch 3, this chapter also suffered from poor inter-observer agreement for the frequency data, while the inter-observer agreement for the duration data was adequate. It is likely that this was due to the same oversight in the configuration outlined in Ch 3. Unfortunately, this error was not discovered in time to correct it for this chapter.

While evidence suggests that the provision of EE can reduce stress in captive cats (especially a hiding box), and that bold and shy cats experience the stress of caging differently, the fact that there was no significant interaction between treatment group and bold/shy behavioural style in any of the most easily interpreted indictors of stress (i.e. any of the variables related to eating, FGM, or either of the CSS variables) suggests that when the provision of EE has shown to be beneficial for cats, it is beneficial regardless of whether the cats was bold or shy. Likewise, when a potential indicator of stress (i.e. any of the variables related to eating, FGM, or either of the CSS variables) has shown a significant difference between bold and shy cats, there was no interaction with treatment group. This means that despite the supposition that bold and shy cats may benefit differently from different types of EE, at least in this experiment this was not the case; both bold and shy cats exhibited lower stress when provided with the hiding box, but not when provided with the shelf.

7.5.1 Conclusions

Many of the strongest potential indicators of stress (but not all) changed significantly over time, giving evidence that for cats, cage confinement is a stressor that diminishes over time, potentially stabilising after 4-6 days. Whether a cat was bold or shy influenced the expression of stress. Bold cats had a lower probability of registering a CSS≥3, and bold cats in BOX spent a significantly lower percentage of their time in the hiding box than did shy cats in BOX. No obvious indicator of stress had a significant interaction between treatment group and whether a cat was bold or shy, indicating that bold cats did not benefit from one type of enrichment while shy cats benefited from another. Finally, different types of EE influenced the expression of stress in singly housed cats. Cats in BOX registered significantly lower FGM concentrations and ate significantly more food per day than did cats in CTRL, while cats in SHELF were not significantly different from either. This evidence suggests that provision of a hiding box may be an easy and important strategy for helping cats cope with the stressors implicit in singly housed confinement.

7.6 References

Carlstead, K., Brown, J.L., Strawn, W., 1993. Behavioral and Physiological Correlates of Stress in Laboratory Cats. Appl. Anim. Behav. Sci. 38, 143-158.

Casey, R.A., Bradshaw, J.W., 2005. The assessment of welfare, in: Rochlitz, I. (Ed.), The Welfare of Cats. Springer, Dordecht, The Netherlands, pp. 23-46.

Cooper, J.O., Heron, T.E., Heward, W.L., 2007. Applied Behavior Science, 2nd Ed. Pearson, Upper Saddle River, N.J.

Deluca, A.M., Kranda, K.C., 1992. Environmental enrichment in a large animal facility. Lab Anim. 21, 38-44.

Dohoo, I.R., Stryhn, H., Martin, S.W., 2009. Veterinary Epidemiologic Research, 2nd ed. VER, Charlottetown, PE, Canada.

Dybdall, K., Strasser, R., Katz, T., 2007. Behavioral differences between owner surrender and stray domestic cats after entering an animal shelter. Appl. Anim. Behav. Sci. 104, 85-94.

Ferguson, C.J., 2009. An effect size primer: A guide for clinicians and researchers. Professional Psychology: Research and Practice 40, 532-538.

Fraser, D., 2008. Understanding Animal Welfare: The Science in its Cultural Context. Wiley-Blackwell, Ames, Iowa.

Godijn, L., 2013. Will a hiding box provide stress reduction for shelter cats? Masters thesis, Universiteit Utrecht, The Netherlands.

Hanley, J.A., Negassa, A., Forrester, J.E., 2003. Statistical analysis of correlated data using generalized estimating equations: an orientation. Am. J. Epidemiol. 157, 364-375.

Jansen, R., Wiertz, L., Meyer, E., Noldus, L., 2003. Reliability analysis of observational data: Problems, solutions, and software implementation. Behavior Research Methods 35, 391-399.

Kessler, M.R., Turner, D.C., 1999. Effects of density and cage size on stress in domestic cats (*Felis silvestris catus*) housed in animal shelters and boarding catteries. Anim. Welfare 8, 259-267.

Kessler, M.R., Turner, D.C., 1997. Stress and adaptation of cats (*Felis silvestris catus*) housed singly, in pairs and in groups in boarding catteries. Anim. Welfare 6, 243-254.

Kry, K., Casey, R., 2007. The effect of hiding enrichment on stress levels and behaviour of domestic cats (*Felis sylvestris catus*) in a shelter setting and the implications for adoption potential. Anim. Welfare 16, 375-383.

Landis, J.R., Koch, G.G., 1977. The Measurement of Observer Agreement for Categorical Data. Biometrics 33, pp. 159-174.

Martin, P., Bateson, P., 2007. Measuring Behaviour : An Introductory Guide, 3rd ed. Cambridge University Press, Cambridge.

McCune, S., 1995. The Impact of Paternity and Early Socialization on the Development of Cats Behavior to People and Novel Objects. Appl. Anim. Behav. Sci. 45, 109-124.

Möstl, E., Palme, R., 2008. Measuring Faecal Steroid Metabolites with Enzyme Immunoassays (EIA) on Microtitre Plates using Biotinylabled Steroids as Labels. University of Veterinary Medicine, Vienna, Austria, pp. 1-10.

Ottway, D.S., Hawkins, D.M., 2003. Cat housing in rescue shelters: A welfare comparison between communal and discrete-unit housing. Anim. Welfare 12, 173-189.

Palme, R., Möstl, E., 1997. Measurement of cortisol metabolites in faeces of sheep as a parameter of cortisol concentration in blood. Zeitschrift fur Saugetierkund 62 Supp/2, 192-197.

Palme, R., Schatz, S., Möstl, E., 2001. [Influence of a vaccination on faecal cortisol metabolite concentrations in cats and dogs.]. DTW (Deutsche Tieraerztliche Wochenschrift) 108, 23-25.

Podberscek, A.L., Blackshaw, J.K., Beattie, A.W., 1991. The Behavior of Laboratory Colony Cats and their Reactions to a Familiar and Unfamiliar Person. Appl. Anim. Behav. Sci. 31, 119-130.

Smith, D.F., Durman, K.J., Roy, D.B., Bradshaw, J.W., 1994. Behavioural aspects of the welfare of rescued cats. The Journal of the Feline Advisory Bureau 31, 25-28.

Tanaka, A., Wagner, D.C., Kass, P.H., Hurley, K.F., 2012. Associations among weight loss, stress, and upper respiratory tract infection in shelter cats. J. Am. Vet. Med. Assoc. 240, 570-576.

Touma, C., Palme, R., 2005. Measuring fecal glucocorticoid metabolites in mammals and birds: the importance of validation. Ann. N. Y. Acad. Sci. 1046, 54-74.

UK Cat Behaviour Working Group, 1995. An Ethogram for Behavioural Studies of the Domestic Cat (*Felis silvestris catus*). Universities Federation for Animal Welfare, Wheathampstead, Herts.

Wemelsfelder, F., 2005. Animal boredom: understanding the tedium of confined lives, in: McMillan, F.D. (Ed.), Mental Health and Well-Being in Animals. Blackwell Publishing, Oxford, pp. 79-92.

8 EVALUATION OF TWO TESTS TO DISCRIMINATE BETWEEN BOLD AND SHY CATS

8.1 Abstract

Bold and shy cats may benefit from different management practices in a shelter setting. Two behavioural tests – an emergence test (ET) and an approach test (AT) were evaluated for their ability to discriminate between bold and shy cats. Tests were evaluated for their agreement and reliability cross-context and cross-time, agreement with observer ratings, as well as for the percentage of cats they correctly identified as bold or shy, using qualitative observer ratings as the reference test. Eighty-four cats were subjected to the ET and AT three times: once at the shelter, once immediately after arrival in the research room, and once in the research room after 10 days. The ET showed acceptable cross-context consistency but the AT did not. Neither test showed sufficient cross-time consistency. In order to implement different management practices in a shelter, cats must be correctly identified quickly (i.e. the first iteration). The first two iterations of the ET showed sufficient agreement with observer ratings, while only the second iteration of the AT showed sufficient agreement with observer ratings. The first iteration of the ET with a 9 s cut-point correctly identified 78% of cats overall, 80% bold cats, and 77% of shy cats. The 10 s cut-point proposed in Ch 4 correctly identified 77% of cats overall, 80% bold cats, and 74% of shy cats in the first iteration of the ET. The first iteration of the AT correctly identified 70% of cats overall, 94% bold cats, and 37% of shy cats. As shy cats show more behavioural signs of stress, it is likely that they would derive greater benefit from a tailored husbandry program. These results indicate that the ET (with a 9 or 10 s cut-point) is superior to the AT for discriminating between bold and shy cats for the management purposes, because the ET has sufficient cross-context agreement and reliability, sufficient agreement with observer ratings in the first iteration, and correctly identified a higher percentage of cats overall and – perhaps more importantly – shy cats in particular.

8.2 Introduction

There is evidence (Ellis Ch 3, this text; Kessler and Turner, 1997) that the stress associated with shelter settings is felt most severely by cats during the first few days after arrival, and reduces over time in association with habituation. As the differentiation between bold and shy cats is based on their reaction to novelty (McCune, 1995), it is likely that these two types of cats may respond to these conditions differently. Therefore, it could be advantageous for shelter staff to be able to classify cats as either bold or shy immediately, so that resources appropriate for each behavioural style could be allocated.

In a shelter environment, behavioural style is often assessed using qualitative observer ratings once the animal has been in the shelter long enough for the staff to form an opinion. Unfortunately, this opinion cannot be formed immediately upon the animal's entrance in the shelter, and thus distribution of appropriate resources can be delayed, missing the most stressful period. The easiest way to assess a behavioural style quickly is through a behavioural test that has been previously validated to discriminate between the dichotomous types.

The study of abstract concepts such as temperament has been criticised for the subjectivity of methodology and interpretation of results (Manteca and Deag, 1993). However, often temperament is assessed in a manner that easily lends itself to stringent testing for validity, agreement, and reliability. The importance of this testing is paramount, and unfortunately often overlooked (see Burns, 2008). A framework for this analysis is rooted in the concept of construct validity. This term refers to how well the operational definition of a concept reflects its theoretical meaning, and can be further broken down into sub-categories of validity, including face validity, internal validity, and convergent validity (Gosling, 2001; Miller et al., 2005). Face validity is a subjective assessment of whether the test appears to measure the theoretical concept being described ‘on face value’ (Burns, 2008). Internal validity refers to the agreement and reliability of the test across time and context (Cozby and Bates, 2012). Finally convergent validity refers to how well a measure compares with other measures thought to assess the same concept (Burns, 2008). The results of a study of behavioural styles that is able to stand up to the rigour of these tests of validity would be difficult to refute as subjective.

Ch 4 of this text showed that an emergence test with a 10 s cut-point had good construct validity as a test to dichotomise bold and shy cats. This test was used again in Ch 5 of this text, where it was found to have poor agreement with qualitative observer ratings, in terms of its ability to discriminate between bold and shy cats. Unfortunately, as agreement was the only measure of validity used in Ch 5, a thorough investigation of how the test failed to satisfy the construct validity was impossible. The purpose of this chapter was to evaluate the construct validity of this test to dichotomise bold and shy cats using a larger sample size, as well as to re-evaluate the established 10 s cut-point for the emergence test. Additionally, the potential of a second test to dichotomise bold and shy cats – the approach test – was evaluated with the same focus on construct validity. Latency to approach a person has been used in previous studies of feline behavioural style (Kry and Casey, 2007; McCune, 1992).

8.3 Methods

8.3.1 Subjects

This study was conducted using 84 domestic cats obtained from a local animal shelter, of which 38 were males (28 were intact and 10 were neutered) and 46 were females of unknown reproductive status. Fifty-five cats came into the shelter as strays and 29 cats were surrendered by their owner. While determining exact ages of shelter cats can be difficult, all of the cats were judged to be adults by the absence of kitten teeth. Thirty-nine of the cats were judged to be 2 years of age or younger, while 45 of the cats were judged to be over 2 years of age. The average amount of time at the shelter before transported to the study site approximately 5 km away was 5.0 ± 2.49 (±SD) days. As a result of recording equipment malfunction, 12 of the cats included in this study could not be used in Ch 7. All cats were deemed to be in reasonably good health by shelter staff before inclusion in the study, and were given physical veterinary examinations upon arrival at the University and again before return to the shelter after the study's completion. The study commenced on September 13th, 2011 and concluded on May 24th, 2012. At any one time there was between 3 and 6 cats in the cages.

8.3.2 Housing and management

An additional study was carried out on a subset of these cats in the previous chapter. All cats were housed in banks of cage, and were in the study for 10 days. For detailed housing and management information on this study, please refer to Ch 7, this text.

8.3.3 The Emergence Test

The first test evaluated for its ability to discriminate between bold and shy cats was the Emergence Test (ET). This test involved confining the cat to a small Kennel Cab II cat carrier (Petmate, Arlington, Texas, USA) (48.3 cm x 31.8 cm x 25.4 cm), placing the carrier in an unfamiliar environment, and measuring the latency to emerge (≥50% of the body) to a maximum of 30 s. Any cat that did not emerge within the test period was assigned a maximum score of 30 s. This test was shown in a preliminary experiment to reliably discriminate between bold and shy cats, at a cut point of 10 s (Ellis Ch 4, this text). However, less consistent results were obtained using this result in a follow-up study (Ellis Ch 5, this text). This investigation with a larger sample size was conducted to help elucidate the reliability of the ET. A short latency to emerge has been associated with boldness in many studies (e.g., Brown et al., 2005; Harcourt et al., 2009). Furthermore, the emergence test in a novel environment is comparable to McCune's Novel Box Test (1995), which used emergence as one of its measures: McCune (1995) found emergence to be a key determinant of 'boldness'. Finally, considering McCune's definition of the bold/shy behavioural style – general response to novelty irrespective of whether the novelty is human or object – it is clear that on face value the ET would be a good test of boldness. The ET was administered once at the local Humane Society (ET iteration 1) using a small Kennel Cab II cat carrier (48.3 cm x 31.8 cm x 25.4 cm) in a closed office environment, and twice at the study location (ET iteration 2 on day one, and ET iteration 3 on day 10) in an empty open field measuring 1.22 m tall with a circumference of 9.75 m, an area of 7.56 m^2, and the same cat carrier protruding from the side and opening into the arena.

8.3.4 The Approach Test

The second test evaluated for its ability to discriminate between bold and shy cats was the Approach Test (AT). In this test, the researcher stood 10 cm in front of the closed cage doors for 60 s. The reaction of the cat was recorded as 'approach', 'retreat', 'no reaction A' (with the cat at the back of the cage, assuming `stressed posture`) or no 'reaction B' (location other than at the back of the cage, posture `seemingly relaxed`) (adapted from Kry and Casey, 2007). Assessments of 'stressed posture' or 'seemingly relaxed' were subjective judgments made by the researcher, but for general guidelines the behavioural definitions given in the Cat-Stress-Score matrix created by Kessler and Turner (1997) were used, focussing most closely on body, head, and tail positions, as well as size of pupils and presence/nature of vocalizations. As Kessler and Turner (1999) identified scores of up to 3 as `acceptable levels` of stress, the exhibition of postural and behavioural elements of the matrix associated with scores greater than 3 were taken as evidence of ‘assuming a stressed posture’, while the absence of these elements was taken as evidence of being ‘seemingly relaxed’. In cases where the cat’s response was recorded as ‘approach’, the latency to approach was also recorded. The AT was administered once at the local Humane Society (AT iteration 1) in the receiving room, and once daily at the study location (the first day’s was considered AT iteration 2, and the final day’s was considered AT iteration 3) in their banks of cages. Although some habituation to the AT may have occurred – as it was administered daily – it was conducted at the same time as the Cat-Stress-Score (see Ch 7, this text), which was already being conducted daily.

8.3.5 Qualitative observer ratings

On their last day in the study (day 10), the investigator assigned each cat a designation as either bold or shy according to McCune’s definition based on her experience with them over their 10 days in the study. For purposes of discrimination, anecdotal observation of a generally positive or neutral response to novelty was interpreted as a bold cat, while a generally negative response to novelty was interpreted as a shy cat.

8.3.6 Procedure

Once every 2 weeks the researcher went to the shelter to pick up a new batch of cats. All cats in the receiving room that were deemed to be adult, and met appropriate health standards, were included in the study. If more than 6 cats met the inclusion criteria, cats were excluded at random. At the shelter after subject selection was complete, AT iteration 1 was conducted. Cats were placed into carriers and taken to an office to conduct ET iteration 1. Afterwards cats were transported to the research facility. Upon arrival, ET iteration 2 was conducted. Cats were then given a veterinary examination and placed into their cages. Starting the following day, additional ATs were conducted at 11:00 h for the next 9 days, but analysis of these tests was not conducted in this chapter. On the final day, AT iteration 3 was conducted at 13:00 h. The cats were then assigned a qualitative observer rating by the researcher as either bold or shy, given a second veterinary examination, and returned to the shelter.

8.3.7 Statistics

Descriptive statistics

The median and interquartile range (IQR) were calculated for latency to emerge in the ET, and for latency to approach in the AT for all three iterations of each test. Additionally, the numbers of individuals identified as bold or shy using the tests were also calculated for each iteration. The 10 s cut-point identified by Ellis (Ch 4, this text) was used to dichotomise cats in the ET. In the AT, cats were designated as either bold or shy based on the cat's categorical reaction to approach: approach and no reaction B were considered indicative of a bold cat, while retreat or no reaction A were considered indicative of a shy cat.

Internal validity

Cross-context consistency

To evaluate cross-context consistency in either test, iteration 1 was compared to iteration 2, looking at the data in a number of formats. For the ET, dichotomous classifications were compared for agreement with Cohen's kappa, using the 10 s cut-point identified by Ellis (Ch 4, this text) to classify the cats as either bold or shy. For the AT, categorical AT responses (described above) were compared for agreement using Fleiss's kappa, and the subsequent dichotomisation of the categorical responses (also described above) were compared for agreement using Cohen's kappa. Analysis of rank agreement was conducted with a Spearman's correlation coefficient. Comparison for agreement of raw numbers was conducted with a concordance correlation coefficient (CCC). Comparison for reliability of raw numbers was assessed using the intraclass correlation coefficient (ICC) (Lessells and Boag, 1987).

Cross-time consistency

To evaluate cross-time consistency in either test, iteration 2 was compared to iteration 3, by examining the data using the same formats and statistical methods described for cross-context comparison.

Convergent validity

The dichotomous outcomes of both the ET and the AT were compared to the dichotomous outcome of the qualitative observer ratings using Cohen's kappa. Additionally, the cut-point for dichotomising bold and shy for the ET set by Ellis (Ch 4, this text) was reassessed. Using the qualitative observer ratings as the reference test, calculations were conducted to determine the percentage of cats correctly identified over all, as well as for bold and shy cats separately, for all of the latencies observed in each of the three iterations of both the ET and the AT. For the ET, this was conducted for the cut-points with the highest percentage of cats correctly classified overall in each iteration, as well as using the 10 s cut-point within each iteration.

Results were analysed for agreement and reliability (de Vet, 2005). Agreement was measured using kappa, Spearman's, and concordance correlation coefficients (CCC), while reliability was assessed using intraclass correlation coefficients (ICC). Minimum acceptability of results was set at >0.6 (substantial or better) for the kappa test statistics (Landis and Koch, 1977), ≥0.5 (a large effect size) for Spearman's correlation coefficient and CCC (Cohen, 1988), and ≥0.5 (indicating consistent individual responses) for ICC (Gibbons et al., 2009). All analyses were conducted using Minitab® 15 and STATA® 10 statistical software.

8.3.8 Ethical approval

This project was approved by UPEI's Animal Care Committee under protocol number 09-051, and followed the guidelines of the Canadian Council of Animal Care's "Guide to the Care and Use of Experimental Animals".

8.4 Results

8.4.1 Descriptive statistics

The median and interquartile range (IQR) for latency to emerge in the ET, and latency to approach in the AT, as well as the numbers of bold and shy cats identified by each test can be seen in Table 8.1.

Table 8.1 Median and IQR of latencies (s) to either emerge or approach in the ET and AT respectively, as well as the number of cats identified as either bold or shy by each test (N=84)

	ET			AT		
	Latency to emerge (s)	No. of cats classified		Latency to approach (s)	No. of cats classified	
	Median (IQR)	Bold	Shy	Median (IQR)	Bold	Shy
iteration 1	8 (19.3)	47	37	7 (58)	68	16
iteration 2	7 (27.8)	50	34	60 (54)	54	30
iteration 3	3 (4.8)	68	16	11 (55.3)	79	5

8.4.2 Internal validity

Cross-context consistency

The ET achieved acceptable levels of cross-context consistency in all three formats. The AT showed close to substantial cross-context agreement only as a dichotomous variable, but did achieve sufficient cross-context reliability in the ICC (Table 8.2).

Table 8.2 Cross-context and cross-time consistency of the ET and AT

	Format of data:	Dichotomous	Categorical	Ranks	Continuous (ln transformed)	
Consistency	Parameter:	Cohen's kappa	Fleiss's kappa	Spearman's correlation	CCC	ICC
Cross-context	ET	0.63, p<0.01	N/A	0.65, p<0.01	0.62, p<0.01	0.62
	AT	0.59, p<0.01	0.46, p<0.01	0.39, p<0.01	0.45, p<0.01	0.50
Cross-time	ET	0.19, p=0.02	N/A	0.39, p<0.01	0.31, p<0.01	0.38
	AT	0.20, p<0.01	0.18, p<0.01	0.24, p=0.03	0.21, p=0.03	0.24

ln=natural log

Cross-time consistency

Both the ET and the AT failed to achieve sufficient cross-time consistency in any of the formats (Table 8.2).

8.4.3 Convergent validity

The first two iterations of the ET showed substantial agreement (or at least approached substantial agreement) with the observer ratings, while the third iteration showed only slight agreement. The AT approached substantial agreement with the observer ratings for the second iteration, but only fair agreement for the first iteration and slight agreement for the third iteration (Table 8.3).

For the reassessment of the cut-point for the ET, the percentage of cats correctly identified overall, as well as for bold and shy cats separately for each potential cut-point for each iteration of the test can be found in Table 8.4. Latencies tested as potential cut-points were those with the highest percentage of cats correctly identified overall within each iteration, as well as using the 10 s cut-point identified by Ellis (Ch 4, this text).

The percentage of bold and shy cats correctly identified by the dichotomised outcome of each iteration of the AT can be found in Table 8.5.

Table 8.3 Results of Cohen's kappa for agreement between dichotomous classification of either ET or AT and the qualitative observer rating

	Iteration 1	Iteration 2	Iteration 3
ET	0.55, p<0.01	0.61, p<0.01	0.09, p=0.09
AT	0.34, p<0.01	0.58, p<0.01	0.05, p=0.47

Table 8.4 Percentage of cats correctly classified (when compared against qualitative observer ratings) for the ET using different potential cut-points

	Potential cut-point (s)	Percentage of cats correctly classified		
		Overall	Bold	Shy
Iteration 1	9	78	80	77
	10†	77	80	74
Iteration 2	15	83	92	71
	25	83	94	69
	10†	81	86	74
Iteration 3	5	62	82	34
	6	62	84	31
	10†	60	86	23

† Cut-point was identified in Ellis Ch 4, this text

Table 8.5 Percentage of both bold and shy cats correctly classified by the dichotomised outcome of the AT when compared against qualitative observer ratings

	Percentage of cats correctly classified		
	Overall	Bold	Shy
Iteration 1	70.2	94	37
Iteration 2	79.8	88	69
Iteration 3	59.5	96	9

8.5 Discussion

Neither the ET nor the AT met all of the standards established for construct validity. While the results of each test were analysed in multiple formats, the validity of the dichotomous outcome variable is the most important variable, as the purpose of this test is to quickly dichotomise between bold and shy cats. The ET showed acceptable levels of cross-context consistency using the dichotomised outcome, and the AT was just outside of the established level of acceptability. However, while the ET showed acceptable cross-context consistency in every format, the AT failed to show acceptable levels in any of the formats investigated.

Furthermore, neither the ET nor the AT showed acceptable levels of cross-time consistency in any of the formats. It is possible that neither of these tests reflected individual differences consistently expressed across time. It is also possible that the responses to the third iteration (especially of the ET) were simply muted to a degree that differences between individuals were imperceptible. Table 8.1 shows that median values of latency to emerge were similar from iteration 1 to iteration 2, with large IQRs. However, iteration 3 had a much lower median value, and a much smaller IQR. It is likely that responses to the first two iterations were heightened due to the stress of recent transport and the novelty of surroundings, while the third iteration of the test was conducted in an environment now familiar with a stable routine, and so fear was reduced and differences between individuals were muted. Behavioural consistency increases in stressful situations (Budaev, 1997). The extremely low median and IQR suggest that response variance may have been muted to almost an undetectable degree, disguising individual differences in any of the data formats, and thus reducing the reliability as expressed by the ICC. However, it is also possible that the cats were becoming habituated to the carrier. The same pattern was not observed for the AT, suggesting that different processes were at work.

Traditional research into temperament traits/behavioural styles suggests that while the magnitude of individual responses can change over time, the rank-order consistency of individuals in a group could stay relatively stable (Roberts and DelVecchio, 2000) although, concentrations of extreme values of a test can result in little variability between individuals and influence rankings of individuals. While neither test achieved a large effect size in the Spearman's analysis of cross-time agreement, the ET achieved a medium effect size, while the AT achieved only a small effect size (Cohen, 1988). Furthermore, many studies of temperament traits/behavioural styles consider any positive correlations as evidence of agreement, provided they were significant (e.g. Visser et al., 2001), although this interpretation is not very restrictive.

In terms of convergent validity, ideally all iterations of tests would achieve acceptable agreement with the reference test. However, since the purpose of this study was to identify a test that can reliably dichotomise individuals immediately upon entry into the shelter, agreement would be most important for the first iteration of this test – since they are in the shelter environment, and have not been there long, or the second iteration of this test – since they have just undergone a stressful event that may mimic initial arrival at the shelter. Both tests were judged to at least approach acceptable agreement on the second iteration of the test, only the ET showed close to acceptable agreement on the first iteration, and neither test approached acceptable agreement on the third iteration. Therefore, the ET can be considered to have shown better convergent validity. It makes sense that the final iteration of the ET had only slight agreement with the reference test, while the first two iterations had moderate agreement, since the first two iterations had substantial agreement with each other, but only slight agreement with the third. However, for the AT, while the first two iterations had moderate agreement with each other, only the second iteration had moderate agreement with the reference test. This indicates that what iterations 1 and 2 had in common with each other were different factors than what iteration 2 had in common with the reference test, signifying that habituation is not the only factor influencing the lack of agreement between the third iteration and the reference test.

The 10 s cut-point held up under investigation of the most appropriate place to dichotomise between bold and shy cats in the ET. It identified almost 75% of shy individuals in both iteration 1 and 2, while bold cats were correctly identified almost 75% and over 85% of the time in these iterations respectively. Admittedly, the 9 s cut-point correctly identified more shy cats in the first iteration, and an identical number of shy cats in the second iteration, and arguably shy cats could benefit from environmental enrichment and tailored husbandry programs more than bold cats (Ellis Ch 4, this text). While this could be an argument for using the 9 s cut point over the 10 s one, a round number such as 10 s may be more intuitive for shelter workers to use, and thus more reasonable to suggest. However, further investigation of the most appropriate cut-point is still necessary – especially in real-world shelter conditions – and at this point there is little reason to suggest either the 9 s or 10 s cut-point over the other.

One major problem with both the ET and the AT in the current design was that while the cross-time consistency assessments were intended to investigate the similarity between two iterations of a test conducted in the same context, this is not exactly what took place. Both iterations were conducted in the same facility, but as one was soon after arrival and one was after 9/10 days of habituation, it could be argued that the cross-time consistency assessments were really simply an additional cross-context consistency assessment. For the purposes of cross-time consistency assessment, ideally two iterations of a test should have been conducted after habituation to the environment and protocols were more or less complete. Unfortunately, the confines of time and money would have made this difficult, and, as this test was designed to be used in a situation in which cats have not yet habituated to the environment, it was potentially not appropriate.

Another potential criticism of the current study design is that the qualitative observer ratings were conducted by only one observer. Ideally to produce qualitative observer ratings, two or more observers would independently rate individuals on behaviourally defined characteristics, and these ratings could be compared for agreement and reliability (Mendl and Harcourt, 2000). In the current study this was not possible due to limited personnel. Although inter-observer agreement and reliability of qualitative observer ratings were not calculated here, ratings of the bold/shy behavioural style produced by the observer in the current study have shown acceptable levels of agreement and reliability in a visual analogue scale, and through dichotomous classification in a previous study (Ellis Ch 4, this text), and this suggests that the observer ratings produced here may be dependable.

Finally, this paradigm implies that the behavioural style bold/shy is dichotomous in cats, as has been suggested by previous research (Meier and Turner, 1985; Mertens and Turner, 1989). In Ellis (Ch 4, this text) the natural log of latency to emerge was observed to express a relatively normal distribution of scores, suggesting that perhaps this behavioural style does exist on a continuum in cats, as has been described in other species (Wilson et al., 1994). In the current study, the fact that both the continuous and rank formatted data showed cross-context and cross-time consistency values similar to those of the dichotomous formatted data suggest that this behavioural style may indeed be continuous. However, there is no indication that dichotomising individuals at a spot in the continuum produces a variable of no value, especially if future studies show that specific types of environmental enrichment or husbandry techniques benefit cats defined as either bold or shy more profoundly. In fact, the dichotomisation of this behavioural style – however unrepresentative of the data set – could have valuable real-world uses.

8.5.1 Conclusion

The ET achieved acceptable levels of cross-context consistency in all formats investigated, whereas the AT failed to reach acceptable consistency in any format. Although neither test achieved acceptable levels of cross-time consistency in any format investigated, the ET almost always outperformed the AT whenever formats were comparable, and descriptive statistics showed that low variance may have accounted for the ET's cross-time consistency failing to reach minimum standards. Additionally, the ET approached acceptable convergent agreement in iterations 1 and 2, while the AT approached acceptable convergent agreement only in iteration 2. Finally, no iteration of the AT was able to correctly identify more than 70% of shy cats, which was previously identified as the more difficult type to identify, and potentially the bigger beneficiary of a tailored husbandry program (Ch 4, this text). The 10 s cut-point identified by Ellis for use with the ET (Ch 4, this text) was able to correctly identity almost 75% of shy cats in both of the first two iterations, while correctly identifying even higher levels of bold cats. For these reasons, the ET was deemed a more consistent and appropriate test than the AT for dichotomising between bold and shy cats. Additionally, the 10 s cut-point previously identified arguably outperformed any of the cut-points identified in the current study, and the recommendation for its use has been upheld. However, context-specific investigations to determine the most appropriate specific cut-points to use in this method are still merited.

8.6 References

Brown, C., Jones, F., Braithwaite, V., 2005. In situ examination of boldness–shyness traits in the tropical poeciliid, *Brachyraphis episcopi*. Anim. Behav. 70, 1003-1009.

Budaev, S.V., 1997. 'Personality' in the guppy (*Poecilia reticulata*): A correlational study of exploratory behavior and social tendency. Journal of Comparative Psychology 111, 399-411.

Burns, J.G., 2008. The Validity of Three Tests of Temperament in Guppies (*Poecilia reticulata*). Journal of Comparative Psychology 122, 344-356.

Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Lawrence Erlbaum Associates, New Jersey.

Cozby, P., Bates, S., 2012. Methods in Behavioral Research, 11th ed. McGraw-Hill, New York.

de Vet, H., 2005. Observer Reliability and Agreement. Encyclopedia of Biostatistics , 1-5.

Gibbons, J., Lawrence, A., Haskell, M., 2009. Responsiveness of dairy cows to human approach and novel stimuli. Appl. Anim. Behav. Sci. 116, 163-173.

Gosling, S.D., 2001. From mice to men: What can we learn about personality from animal research? Psychol. Bull. 127, 45-86.

Harcourt, J.L., Sweetman, G., Johnstone, R.A., Manica, A., 2009. Personality counts: the effect of boldness on shoal choice in three-spined sticklebacks. Anim. Behav. 77, 1501-1505.

Kessler, M.R., Turner, D.C., 1999. Effects of density and cage size on stress in domestic cats (*Felis silvestris catus*) housed in animal shelters and boarding catteries. Anim. Welfare 8, 259-267.

Kessler, M.R., Turner, D.C., 1997. Stress and adaptation of cats (*Felis silvestris catus*) housed singly, in pairs and in groups in boarding catteries. Anim. Welfare 6, 243-254.

Kry, K., Casey, R., 2007. The effect of hiding enrichment on stress levels and behaviour of domestic cats (*Felis sylvestris catus*) in a shelter setting and the implications for adoption potential. Anim. Welfare 16, 375-383.

Landis, J.R., Koch, G.G., 1977. The Measurement of Observer Agreement for Categorical Data. Biometrics 33, pp. 159-174.

Lessells, C.M., Boag, P.T., 1987. Unrepeatable Repeatabilities: A Common Mistake. Auk 104, 116-121.

Manteca, X., Deag, J.M., 1993. Individual differences in temperament of domestic animals: a review of methodology. Anim. Welfare 2, 247-268.

McCune, S., 1995. The Impact of Paternity and Early Socialization on the Development of Cats Behavior to People and Novel Objects. Appl. Anim. Behav. Sci. 45, 109-124.

McCune, S., 1992. Temperament and the welfare of caged cats. Ph.D. Thesis, University of Cambridge.

Meier, M., Turner, D.C., 1985. Reactions of house cats during encounters with a strange person: Evidence for two personality types. Journal of the Delta Society 2, 45-53.

Mendl, M., Harcourt, R., 2000. Individuality in the cat: origins, development and stability, in: Turner, D.C., Bateson, P. (Eds.), The Domestic Cat: the Biology of its Behaviour, 2nd ed. Cambridge University Press, Cambridge, pp. 47-64.

Mertens, C., Turner, D.C., 1989. Experimental analysis of human-cat interactions during first encounters. Anthrozoös 2, 83-97.

Miller, K.A., Garner, J.P., Mench, J.A., 2005. The test-retest reliability of four behavioural tests of fearfulness for quail: a critical evaluation. Appl. Anim. Behav. Sci. 92, 113-127.

Réale, D., Reader, S.M., Sol, D., McDougall, P.T., Dingemanse, N.J., 2007. Integrating animal temperament within ecology and evolution. Biol. Rev. 82, 291-318.

Roberts, B.W., DelVecchio, W.F., 2000. The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. Psychol. Bull. 126, 3-25.

Visser, E.K., van Reenen, C.G., Hopster, H., Schilder, M.B.H., Knaap, J.H., Barneveld, A., Blokhuis, H.J., 2001. Quantifying aspects of young horses' temperament: consistency of behavioural variables. Appl. Anim. Behav. Sci. 74, 241-258.

Wilson, D.S., Clark, A.B., Coleman, K., Dearstyne, T., 1994. Shyness and Boldness in Humans and Other Animals. Trends in Ecology & Evolution 9, 442-446.

9 GENERAL DISCUSSION

The objective of this thesis was to evaluate whether environmental enrichment can attenuate stress due to caging in domestic cats, and to examine the role of the cat's behavioural style in this relationship. Through a series of experiments, evidence was presented that providing singly housed shelter cats with a hiding box can result in significantly lower stress compared to cats provided with no additional cage enrichment, and that although bold and shy cats can be reliably identified – and shy cats express greater signs of stress than do bold – there was no evidence that the two modes of this behavioural style benefited differentially from provision of the selected environmental enrichment devices.

This study has presented some novel and interesting findings beyond the main conclusion. In both Ch 3 and Ch 7, percentage of time spent eating was found to increase, while percentage of time spent grooming was found to decrease with habituation to the environment. This suggests that in future studies these two behaviours could be monitored as potential indicators of stress. Time-windows during which these behaviours were most likely to be initiated were identified (grooming – 23:00-03:00 h and eating – 07:00-11:00 h), thus allowing researchers to focus on recordings from specific parts of the day. In the literature, there exists published time budgets for free-ranging cats (Watanabe et al., 2005), and cats in the home environment (Barry and Crowell-Davis, 1999), and frequency and/or duration data on the behaviours exhibited by communally housed caged cats (Gouveia et al., 2011; Iki et al., 2011). However, Ch 3 presented the first post-habituation time budget for singly housed caged cats which may be useful for cross-study comparison in future studies of similarly housed cats. Ch 4 produced a promising test to discriminate between bold and shy cats in a shelter setting (the emergence test – later supported by the results of Ch 8). Although the literature on the assessment of temperament and behavioural style is plentiful on the domestic dog (summarised in Jones and Gosling, 2005), this work contributes significantly to the relative lack of what the Association of Shelter Veterinarians call “criteria for systematic behavioural assessments in cats” (Newbury et al., 2010). Ch 5 provided the first investigation of visit frequency and time allocation in proximity to types of environmental enrichment within the cat literature, and was the first study to investigate hiding and perching opportunities separately as environmental enrichment options for cats. The hiding box was used more than no enrichment and a prey-simulating toy, but usage of the perching opportunity was not different than that of any of the other environmental enrichment options. Assessment of the agreement and reliability of a Hobo data logger and a proposed data cleaning method to monitor movement through cat flaps in Ch 6 made electronic monitoring of environmental enrichment usage data in Ch 5 possible. In Ch 7, cats identified as shy had significantly greater Cat-Stress-Scores (CSS) than did their bold counterparts when housed singly, regardless of their environmental enrichment treatment group, suggesting that shy cats might suffer more in shelter conditions and could potentially benefit from extra attention. This chapter also provided the first evidence that providing a cat with a hiding box (with no perching opportunity) results in lower FGM concentrations, and higher daily food intake, when compared with cats with no additional environmental enrichment. These findings were independent of whether the cat was bold or shy.

9.1 Limitations

There were some limitations to the study that should be recognised when interpreting any results, and when considering the conclusions. Perhaps the most notable limitation is that a choice test was used to look at the environmental enrichment usage of cats, and not a consumer-demand theory motivation test. Although a motivation test would have allowed investigation of the relative importance of this usage, as mentioned in Ch 5 this approach was abandoned due to the amount of time it took to train cats to use the flap doors, and the interest in evaluating whether usage immediately after arrival was different from after habituation. There was no effect of time identified, but there was a gap between arrival and the start of the experimental period due to training time even without weighting the doors. It is still possible that compartment usage during this initial period could have differed from the fully trained period. However, the most used type of environmental enrichment in Ch 5 also proved to be the type of environmental enrichment in the treatment group associated with the lowest stress response: the hiding box. This gives substantial support to the assumption that the EE compartment usage was motivated by attempting to reduce stress.

Ch 3 (and to a lesser extent, Ch 4) also suffered from a small sample size. Although caution must be taken when interpreting the findings of these studies, there was evidence for replication of the findings in other studies. It is encouraging that similar findings to both of these studies were produced in Ch 7 and Ch 8 respectively. For example, percentage of time spent eating and grooming were related to time in both Ch 3 and Ch 7, while the 10 s cut-point was found to be an appropriate test for discriminating between bold and shy cats in both Ch 4 and Ch 8.
It is also possible that the conditions used in this study did not sufficiently replicate the shelter environment for results to be generalisable to shelter animals. The room environment may have been more akin to conditions common for laboratory cats than for shelter cats, and it is likely that a shelter environment would be more stressful (due to a more varied routine, and higher traffic of novel people and smells). This is alluded to in Ch 3 when discussing how much lower the magnitude of the CSS were than those reported by Kessler and Turner (1997). However, although the magnitude of the responses was lower, the patterns in the data were very similar

to those observed by Kessler and Turner. It seems that if the environment used in this study was less stressful than that of a true shelter, the results would only be magnified in a shelter setting, perhaps even strengthening the conclusions.

Conversely, another possible criticism of this study was that shelter cats were used instead of laboratory cats, for which there would have been greater control over the cat-level characteristics, as specification of age, sex, etc would have been possible when placing the order. Although this would have allowed for smaller sample sizes and likely resulted in less variance, the results would have been less generalisable to an actual population of shelter cats, which would have a large range of variability in previous experience and other cat-level characteristics. Additionally, a condition of the funding body for the project was that the cats must not be sourced from a laboratory supplier.

Unfortunately, there were also a few instances where the data were not supported by sufficient reliability or agreement, negating analysis that would have added to the project, or subduing conclusions. For example, the inter-observer agreement of frequencies of exhibition of behaviours in Ch 3 and Ch 7 were deemed too low to analyse the data. While many interesting results were derived from analysis of the percentage of time spent engaged in some of the behaviours, it is possible that some interesting and significant patterns were missed because analysis of the frequency data were deemed not possible. As was outlined in Ch 3, it is likely that the inadequate inter-observer agreement was a result of an oversight when programming the configuration of the data collection software. The agreement and reliability between the data collected by the Hobo data logger and the data observed from video analysis in Ch 6 were weaker than was desired. Although the frequency data had acceptable agreement and reliability in all five compartments and the percentage of time spent and median duration of visit data had acceptable agreement and reliability in all experimental compartments (with one exception), it was expected that these parameters would be even higher, since they were observations of the exact same events. However, since they surpassed the minimum standards suggested in the literature, they were considered to be reasonable evidence that the data analysed were of sufficiently high quality. Finally, there was a lack of cross-time agreement and reliability in Ch 8 for the Emergence Test. This is unfortunate because as temperament is thought to be consistent across time, so too should be the results of any test designed to assess it. However, as discussed in that chapter, the results of the final iteration of this test were likely influenced by habituation; when this test was conducted all cats emerged relatively quickly. In retrospect, this iteration was likely testing something different than the others. In the first two iterations the emergence of the cats when placed in a novel environment was tested. In the third iteration, the emergence latency was tested in the same room as the second iteration, which was also the room in which cats were weighed daily. It is likely that as the cats were familiar with this room, the environment was perceived as less of a danger, causing emergence to be quicker in all cats.

9.2 Practical applications

Regardless of the limitations, these results could have practical applications for improving the quality of life for singly housed cats. Firstly, this study supports a strong recommendation for including hiding opportunities in enclosure design for all singly housed cats. There was a significant reduction of FGM concentrations associated with the provision of this resource. As elevated glucocorticoids can result in alterations in physiological functions, such as immune function, high concentrations of FGM may mean that an animal is more susceptible to disease. As disease transmission and management is a huge concern in shelters, providing hiding opportunities to each cat may help play a role in reducing incidents of disease. Additionally, one of the behavioural responses to stress can be decreased appetite, which is a common problem at shelters and can compound the problem of insufficient energy being directed towards immune function. As increased food intake was also associated with access to hiding opportunities, this may help to reduce this common problem in shelter care.

There are risks and benefits to housing cats communally. Due to the previous experiences of each cat, and the risk of disease transmission in the shelter, sometimes it is necessary to house cats singly. The Association of Shelter Veterinarians states that single housing must be provided “for animals who are fearful or aggressive towards other animals, are stressed by the presence of other animals, require individual monitoring, or are ill and require treatment that cannot be provided in group housing”, and that single housing is preferable to group housing “when a shorter stay is anticipated” (Newbury et al., 2010). However, this environment must be properly enriched. Provision of hiding (and perching) opportunities for caged cats are already recommended in shelter environments by the Association of Shelter Veterinarians (Newbury et al., 2010), in catteries by the Canadian Veterinary Medical Association (2009), and in laboratory environments by the Canadian Council on Animal Care (Olfert et al., 1993). However, this is the first scientific evidence to show that hiding boxes are the most used type of environmental enrichment of singly housed cats (when compared to toys, a perching opportunity, and no additional enrichment), and that providing it can increase food intake and decrease FGM.

9.3 Environmental requirements, behavioral needs, coping styles, and evolution

As was briefly described in Ch 5, the strength of the evidence that the provision of hiding opportunities reduced the negative effects of shelter conditions on cats presented throughout this thesis may lead some (e.g. Duncan and Olsson, 2001) to suggest that these hiding opportunities are not environmental enrichment at all, but rather are an environmental requirement to meet the basic needs of this animal in these conditions. This line of thinking relies on the premise that cats experience a negative state such as fear in these conditions, and that cats instinctively perform hiding behaviours in response to this state as a coping mechanism. If so, failing to provide hiding opportunities results in the continuation or worsening of the negative state, while providing these opportunities may alleviate the negative effects by catering to the animal's basic needs. While terminology resulting from this argument (i.e. the hiding opportunity being termed an 'environmental requirement' instead of 'environmental enrichment') may vary from what is presented in the preceding chapters, the conclusions and resulting recommendations are the same: the negative experience endured by singly housed cats in shelter conditions can be reduced by providing hiding opportunities, therefore these opportunities should be provided.

The idea that cats instinctively perform hiding behaviours in response to a negative state such as fear, and that failing to provide these opportunities in potentially fearful circumstances fails to meet their basic needs, invokes consideration of the concept of behavioural needs. To define this concept, Dawkins (1983) made the distinction between proximate needs and ultimate goals. The ultimate goals of an animal are achieved through the behaviours that need to happen to ensure survival and reproduction (i.e. feeding and nest-building), while the proximate needs are based on the feelings the animal has that motivate the behaviours aimed at ensuring the ultimate goals are achieved. In captive conditions, while the ultimate goals of the animal are provided for by human intervention, the animal may still be motivated to perform some of the goal-driven behaviours based on their proximate needs. Dawkins suggests that the behavioural needs of an animal may be identified by measuring how motivated an animal is to perform a behaviour when its corresponding ultimate goal has already been satisfied. Using this framework, is hiding a behavioural need for cats? To answer this question, first the possible purpose of the behaviour in relationship to its ultimate goal must be established: surviving a perceived danger through concealment. It must then be asked, was the ultimate goal satisfied in captivity? It is possible that the answer to this question is different for bold and shy cats, due to the differences in their responses to novelty. In bold cats, it is possible that the conditions were not perceived as dangerous, or at least as less dangerous than as perceived by shy cats. In Ch 7, bold cats had significantly lower CSS (as a linear variable) than did shy cats on days 2, 3, and 4, which could be evidence that they perceived the environment as less dangerous initially (however, there was no difference between bold and shy cats in other potential indicators of stress, such as FGM concentrations). So if it is conceivable that the ultimate goal was satisfied (or at least, irrelevant) for bold cats and not for shy cats, hiding would represent a behavioural need if it is exhibited by both groups, despite the difference in perceived danger. While both cats were observed to use the hiding box, shy cats spent a significantly higher percentage of time engaged in this behaviour. Since access to the hiding box was free in this study, the strength of motivation was not measured and it is impossible to determine how strongly the animals were driven to spend time inside the hiding box, but the fact that animals which arguably perceived the environment as less dangerous used the hiding box less, suggests that hiding is not a behavioural need. Instead perhaps it is a strategy for coping with the stress inherent in the perceived danger of the specific potential stressor.

Coping style is a concept similar to the concept of behavioural styles. It is defined as “a coherent set of behavioural and physiological stress responses which is consistent over time and which is characteristic to a certain group of individuals" (Koolhaas et al., 1999). The major difference is that while there can be any number of behavioural styles identified, only two coping styles have been described (Réale et al., 2007): proactive and reactive (sometimes called active and passive/inactive coping) (Koolhaas et al., 1999). The proactive coping style is characterised by exhibition of Cannon's fight-or flight response (Archer, 1979) and when faced with stressors, these individuals will often exhibit aggression or fleeing behaviours. The reactive coping style has been described as the conservation-withdrawal response (Engel and Schmale, 1972) and when faced with stressors, these individuals will often exhibit immobility and low levels of aggression (Koolhaas et al., 1999). Since these styles are defined as a set of behavioural and physiological responses, studies of this concept measuring only one variable have been criticised for not fulfilling the defining criterion. However, there is evidence that parameters that measure initiative or proactivity (such as latency) are the most successful for discriminating between coping styles (summarised in Koolhaas et al., 1999). It would seem then, that perhaps the emergence test in this thesis could have been testing whether an individual had a proactive or reactive coping style instead of whether an individual was bold or shy. In the emergence test, a cat that does not emerge (or emerges after a long latency) would be described as the reactive (or passive/inactive) coping style. This cat's behaviour could also be described as hiding. This relates to the use of the hiding box in the previous paragraph. Perhaps hiding is an important species-specific part of the reactive coping style in cats. Proactive and reactive coping styles in response to confinement have been discussed in the feline literature before. Reactive copers exhibited behaviours similar to what was seen in cats designated as shy in the current study (McCune, 1992), and have been described as more distressed in confinement and take longer to habituate than do proactive copers (Ellis, 2009). Ellis (2009) outlines that although scientific evidence is lacking, anecdotal evidence suggests that proactive copers may benefit more from environmental enrichment stimulating play, while reactive copers may benefit more from hiding opportunities. In Ch 7 of the current thesis it was shown that both bold and shy cats benefited from hiding opportunities, but perching opportunities were investigated instead of play opportunities since the shelf was found to be such a commonly used location in Ch 3, and was the second most used compartment in Ch 5. Although the bold/shy proactive/reactive comparison seems quite logical based on behavioural exhibition, neuroendocrine evidence suggest that perhaps the similarities are not so clear. Koolhaas et al. (1999) state that "coping styles are not only characterised by differences in behavior, but also by differences in physiology and neuroendocrinology", and reviewed a number of studies that revealed significant differences in these parameters between proactive and reactive copers. In the current thesis, Ch 7 revealed no significant difference between the FGM concentrations of bold and shy cats. Although FGM was the only physiological or endocrine variable measured in the current study – and thus others may have revealed differences between the two groups – it is suggestive that perhaps the emergence test was truly measuring the bold/shy behavioural style – behavioural style has not been shown to be linked to glucocorticoid production (Iki et al., 2011; Siegford et al., 2003) – and not coping style. Regardless of whether the emergence test was measuring behavioural style or coping style, for the welfare of singly housed caged cats it may boil down to a matter of semantics: both would support the recommendation to provide hiding opportunities.

Coping styles have evolved to help individuals deal with their environment (although the same or similar would likely be true for behavioural style as well). Within a few generations in the wild, the benefits afforded by each mode would likely produce genetic selection pressure for the distinct genotypes, resulting in a bimodal distribution. Each mode would have varying success depending on the circumstances of the environment (i.e. stability, social structure, food availability) (Koolhaas et al., 1999). For example, in cats there is some evidence that individuality is linked to coat colour, and in stray/feral cat research orange cats are more dominant and aggressive and, due to population densities, fare better in rural population, and consequently, are therefore found in higher frequencies there (Pontier et al., 1995). However, in domesticated circumstances there would be less selection pressure from the environment, as humans provide for many of the animals basic needs (Koolhaas et al., 1999). This could result in a less bimodal representation of the styles, which could explain the relatively normal distribution of the results from the emergence test found in this thesis, despite the fact that the bold/shy behavioural style has been described as discrete for cats in the past (Meier and Turner, 1985; Mertens and Turner, 1989).

9.4 Recommendations for future work

Although this thesis had some important findings, it also revealed a number of areas in which further research would be beneficial. There is some evidence from Ch 4 and Ch 5 that the impact of social enrichment with humans on shelter cats is significant. In Ch 4 there was very little variability between cats in any of the outcome measures in behavioural tests B and C (the two tests that involved human presence). That is because almost all of the cats engaged the humans as often as possible. In Ch 5, one of the three choice chambers was only accessible by people from one side; the compartment on that side was the perching opportunity. Cats in this choice chamber used this compartment significantly more than cats in the other choice chambers. Arguments presented in that chapter suggest that these cats might have spent more time waiting in that compartment for the opportunity for social enrichment in the form of human interaction. These two findings support the common notion that human interaction could be very important to cats, and that enrichment in this form has the potential to be greatly beneficial to shelter cats.

There is much room for refinement of the emergence test with the 10 s cut-point to discriminate between bold and shy cats. It remains to be seen whether it is practical to expect shelters to have an empty room in which this test could be conducted, and what effects the other purposes for which this room is used (i.e. dog/potential adopters introductions and the associated scents) could have on results. It could be that the actual cut-point used needs to be reassessed and validated on an institution by institution basis. There is also the potential for the method proposed in Ch 4 to be used to investigate different behavioural styles, or with different species, if the end goal is to dichotomise between two modes using one behavioural test.

Finally, it would also be quite worthwhile to conduct a consumer-demand motivation test to assess the relative importance of the differing environmental enrichment usage recorded in this thesis. Since the usage was not significantly different across the test period, the time needed to train the animals to operate increasingly weighted doors and in-door out-door system would be unlikely to affect results further, as was feared in the current study. This analysis could reveal that although the hiding box compartment was used most in this study, when the cost of accessing it became too high, cats reduced their use of it, i.e. elastic demand. Moreover it could also reveal that although the cats used the toy compartment the least of all the enriched compartments, access to that compartment stayed at the same level regardless of cost increase, i.e. inelastic demand. This type of data could have implications on the recommendations made to shelters in terms of environmental enrichment provisions. The consumer-demand motivation test could be conducted with the types of environmental enrichment used in the current study, but also with a range of different types of environmental enrichment, including social. It would be interesting then, to conduct an additional experiment similar to that conducted in Ch 7 using the types of environmental enrichment for which cats showed the highest motivation to access. Additionally, in the current study, due to the amount of time it took to train the cats just to use the cat doors, it was difficult to measure compartment usage immediately after arrival. The activity during this period have the potential to help us understand how to best enrich the environment of animals during this particularly stressful time.

9.5 Conclusions

The results of this thesis suggest that holding cats in singly housed confined conditions is a stressor that diminishes with time. There is evidence that the stressors implicit with these conditions are felt more strongly by shy cats than by bold cats, but that provision of a hiding box can help reduce stress in these cats regardless of their mode of this behavioural style. The behavioural and physiological evidence produced here lend strong support to the recommendation that hiding opportunities be made available to singly housed cats in shelters and in similar conditions elsewhere (e.g. laboratories and catteries).

9.6 References

Archer, J., 1979. Animals Under Stress. Edward Arnold, London.

Barry, K.J., Crowell-Davis, S.L., 1999. Gender differences in the social behavior of the neutered indoor-only domestic cat. Appl. Anim. Behav. Sci. 64, 193-211.

Canadian Veterinary Medical Association, 2009. Code of Practice for Canadian Cattery Operations [cited 2013, Oct 17]. Available from: http://www.canadianveterinarians.net/documents/a-code-of-practice-for-canadian-cattery-operations.

Dawkins, M.S., 1983. Battery hens name their price: Consumer demand theory and the measurement of ethological ‘needs’. Anim. Behav. 31, 1195-1205.

Duncan, I.J.H. and Olsson, I.A.S., 2001. Environmental enrichment: from flawed concept to pseudo-science. Proceedings of the 35th International Congress of the International Society for Applied Ethology, Davis, California, p. 73.

Ellis, S.L.H., 2009. Environmental enrichment: practical strategies for improving feline welfare. J. Feline Med. Surg. 11, 901-912.

Engel, G.L., Schmale, A.H., 1972. Conservation-withdrawal. in: Physiology, emotions and psychosomatic illness, Ciba Foundation Symposium 8. Associated Scientific Publishers, New York, New York, pp.57-87.

Gouveia, K., Magalhães, A., de Sousa, L., 2011. The behaviour of domestic cats in a shelter: Residence time, density and sex ratio. Appl. Anim. Behav. Sci. 130, 53-59.

Iki, T., Ahrens, F., Pasche, K.H., Bartels, A., Erhard, M.H., 2011. Relationships between scores of the feline temperament profile and behavioural and adrenocortical responses to a mild stressor in cats. Appl. Anim. Behav. Sci. 132, 71-80.

Jones, A.C., Gosling, S.D., 2005. Temperament and personality in dogs (*Canis familiaris*): A review and evaluation of past research. Appl. Anim. Behav. Sci. 95, 1-53.

Kessler, M.R., Turner, D.C., 1997. Stress and adaptation of cats (*Felis silvestris catus*) housed singly, in pairs and in groups in boarding catteries. Anim. Welfare 6, 243-254.

Koolhaas, J., Korte, S., De Boer, S., Van Der Vegt, B., Van Reenen, C., Hopster, H., De Jong, I., Ruis, M., Blokhuis, H., 1999. Coping styles in animals: current status in behavior and stress-physiology. Neurosci. Biobehav. Rev. 23, 925-936.

McCune, S., 1992. Temperament and the welfare of caged cats. Ph.D. Thesis, University of Cambridge.

Meier, M., Turner, D.C., 1985. Reactions of house cats during encounters with a strange person: Evidence for two personality types. Journal of the Delta Society 2, 45-53.

Mertens, C., Turner, D.C., 1989. Experimental analysis of human-cat interactions during first encounters. Anthrozoös 2, 83-97.

Newbury, S., Blinn, M.K., Bushby, P.A., Cox, C.B., Dinnage, J.D., Griffin, B., Hurley, K.F., Isaza, N., Jones, W., Miller, L., 2010. Guidelines for Standards of Care in Animal Shelters. Association of Shelter Veterinarians.

Olfert, E.D., Cross, B.M., McWilliam, A.A., 1993. Guide to the Care and use of Experimental Animals. Canadian Council on Animal Care Ottawa.

Pontier, D., Rioux, N., Heizmann, A., 1995. Evidence of Selection on the Orange Allele in the Domestic Cat *Felis catus*: The Role of Social Structure. Oikos 73, pp. 299-308.

Réale, D., Reader, S.M., Sol, D., McDougall, P.T., Dingemanse, N.J., 2007. Integrating animal temperament within ecology and evolution. Biol. Rev. 82, 291-318.

Siegford, J.M., Walshaw, S.O., Brunner, P., Zanella, A.J., 2003. Validation of a temperament test for domestic cats. Anthrozoos 16, 332-351.

Watanabe, S., Izawa, M., Kato, A., Ropert-Coudert, Y., Naito, Y., 2005. A new technique for monitoring the detailed behaviour of terrestrial animals: A case study with the domestic cat. Appl. Anim. Behav. Sci. 94, 117-131.

10 APPENDICES

Appendix A – Chapter 5

Table A.1 Daily median (M) and IQR (I) for frequency of compartment visit (N=25)

		Day since trained						
		1	2	3	4	5	6	7
Compartment	CTRL	M=3 I=4.5	M=2 I=3.5	M=4 I=5.5	M=3 I=5.0	M=3 I=4.5	M=2 I=5.5	M=3 I=3.5
	TOY	M=2 I=4.0	M=2 I=5.2	M=2 I=7.0	M=4 I=7.0	M=2 I=4.5	M=3 I=3.5	M=2 I=3.5
	SHELF	M=4 I=5.5	M=3 I=5.5	M=3 I=5.0	M=4 I=7.0	M=4 I=5.5	M=3 I=5.0	M=3 I=2.5
	BOX	M=5 I=4.5	M=5 I=4.0	M=6 I=5.5	M=5 I=5.0	M=6 I=6.0	M=5 I=4.5	M=5 I=5.5

Table A.2 Daily median (M) and IQR (I) for percentage of time spent in each compartment (N=25)

		Day since trained						
		1	2	3	4	5	6	7
Compartment	CTRL	M=1.0 I=4.39	M=1.3 I=4.66	M=0.8 I=6.58	M=1.8 I=6.61	M=2.5 I=3.65	M=0.8 I=3.87	M=1.0 I=3.52
	TOY	M=0.4 I=1.92	M=0.7 I=7.65	M=1.7 I=16.82	M=2.5 I=12.28	M=1.2 I=12.53	M=2.2 I=4.56	M=1.2 I=7.07
	SHELF	M=9.2 I=33.48	M=4.4 I=45.06	M=3.8 I=38.64	M=2.5 I=45.89	M=6.3 I=35.90	M=5.7 I=31.30	M=4.8 I=43.10
	BOX	M=61.5 I=65.69	M=60.7 I=64.04	M=50.4 I=46.70	M=38.3 I=52.24	M=49.9 I=55.07	M=59.5 I=57.51	M=62.3 I=73.45

Appendix B – Chapter 7

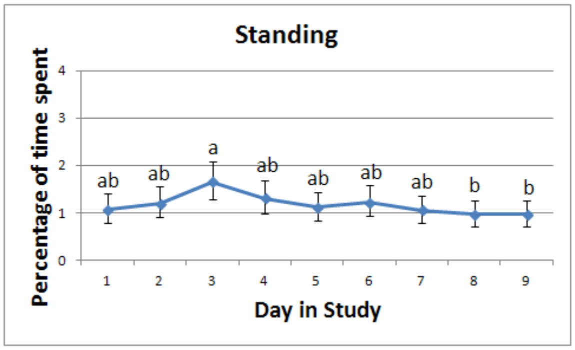

Figure B.1 Effect of day in study on percentage of time spent standing. Graphs are of means, and error bars are confidence intervals. Percentages of time spent eating and grooming has been back-transformed. Results presented for percentage of time spent eating and amount of food eaten daily were Bonferroni-adjusted, results for percentage of time spent grooming were not Bonferroni adjusted. Treatment groups with the same letter code were not significantly different (P>0.05).

a b c

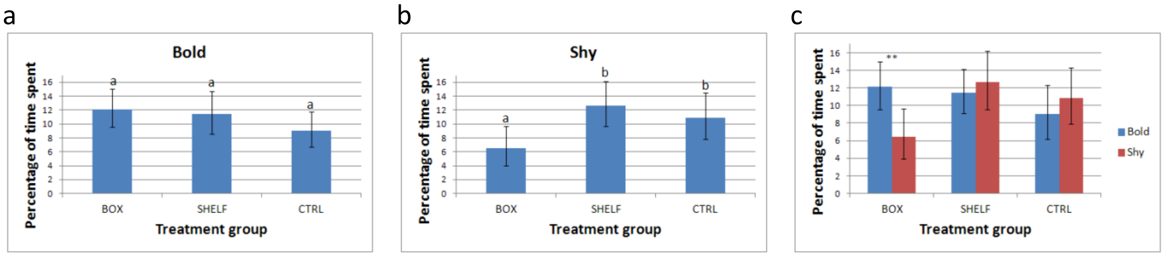

Figure B.2 Combined effects of 'treatment group' and 'mode of behavioural style' on percentage of time spent sitting. Results presented are of back-transformed means, and the error bars are 95% confidence intervals. Figure B.2a shows the differences between treatment groups in bold cats, Figure B.2b shows the differences between treatment groups in shy cats, and Figure B.2c shows the differences between bold and shy cats within treatment groups. In Figures B.2a and B.2b, treatment groups with the same letter code were not significantly different (P>0.05), while Figure B.2c uses asterisks to indicate significant difference between bold and shy cats. Two asterisks mean P<0.01.

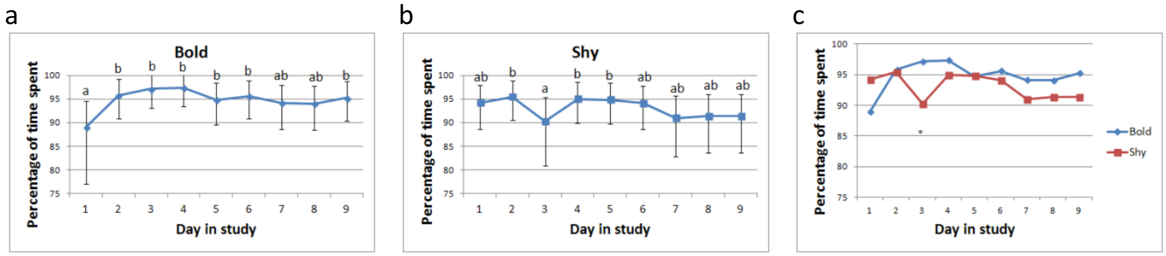

Figure B.3 Combined effects of 'day in study' and 'mode of behavioural style' on percentage of time spent on the cage floor (for SHELF cats only). Results presented are of back-transformed means, and the error bars are 95% confidence intervals. Figure B.3a presents the changes in bold cats over time, Figure B.3b shows the changes in shy cats over time; treatment groups with the same letter code were not significantly different (P>0.05). Figure B.3c shows the differences between bold and shy cats over time. An asterisks indicates a significant difference between bold and shy cats on a particular day. One asterisk means P<0.05.

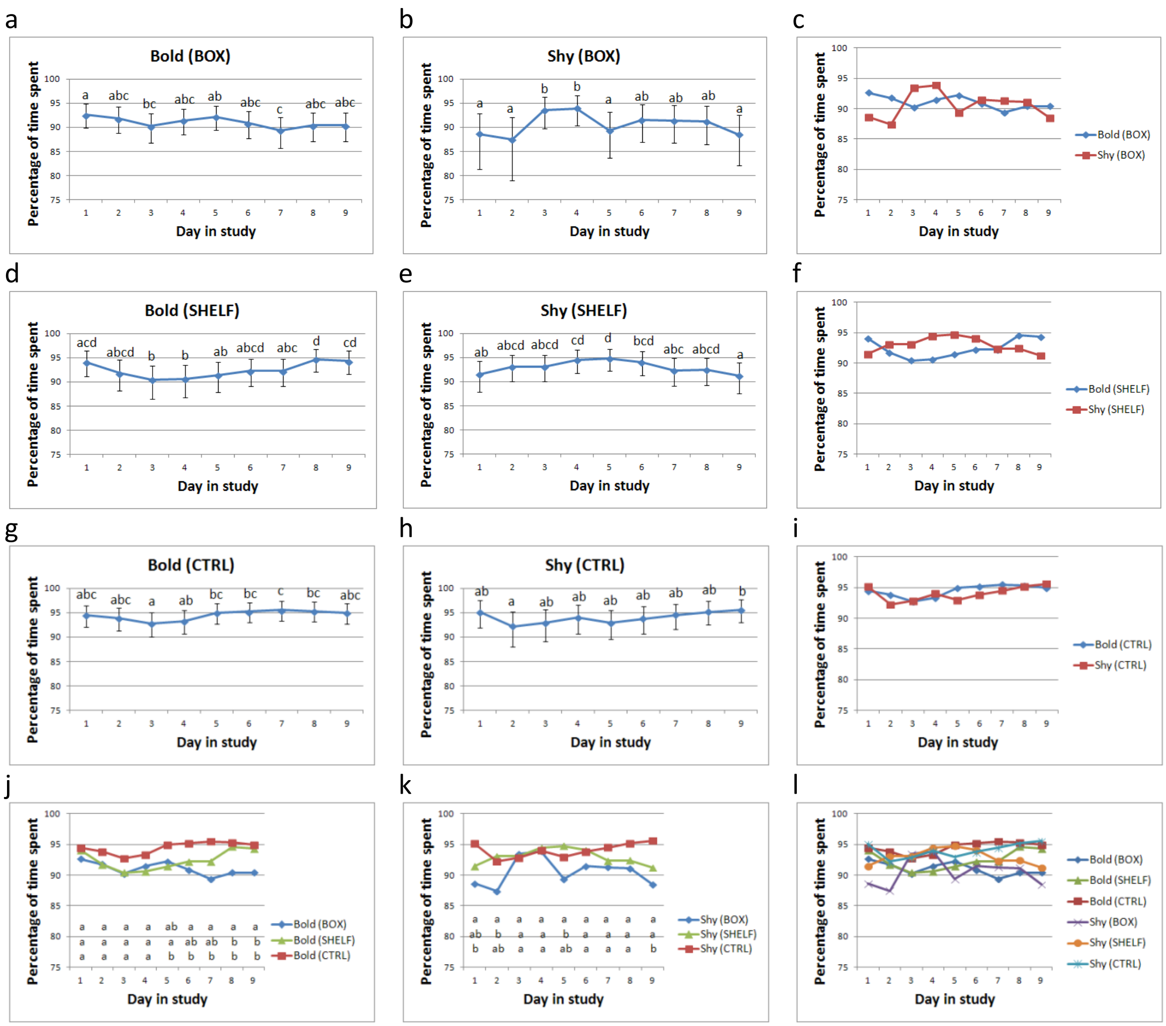

Figure B.4 Combined effects of 'day in study', 'mode of behavioural style' and 'treatment group' on percentage of time spent resting. Results presented are of back-transformed means, and the error bars are 95% confidence intervals. Figures B.4a-c present BOX cats, Figures B.4d-f present SHELF cats, Figures B.4g-i present CONTROL cats, and Figures B.4j-l present all treatment groups together. The first column of figures in each row shows the changes in bold cats over time, and the second column shows the changes in shy cats over time; treatment groups with the same letter code were not significantly different (P>0.05). The final column shows the differences between bold and shy cats over time. These figures use asterisks to indicate a significant difference between bold and shy cats on particular days (P<0.05).

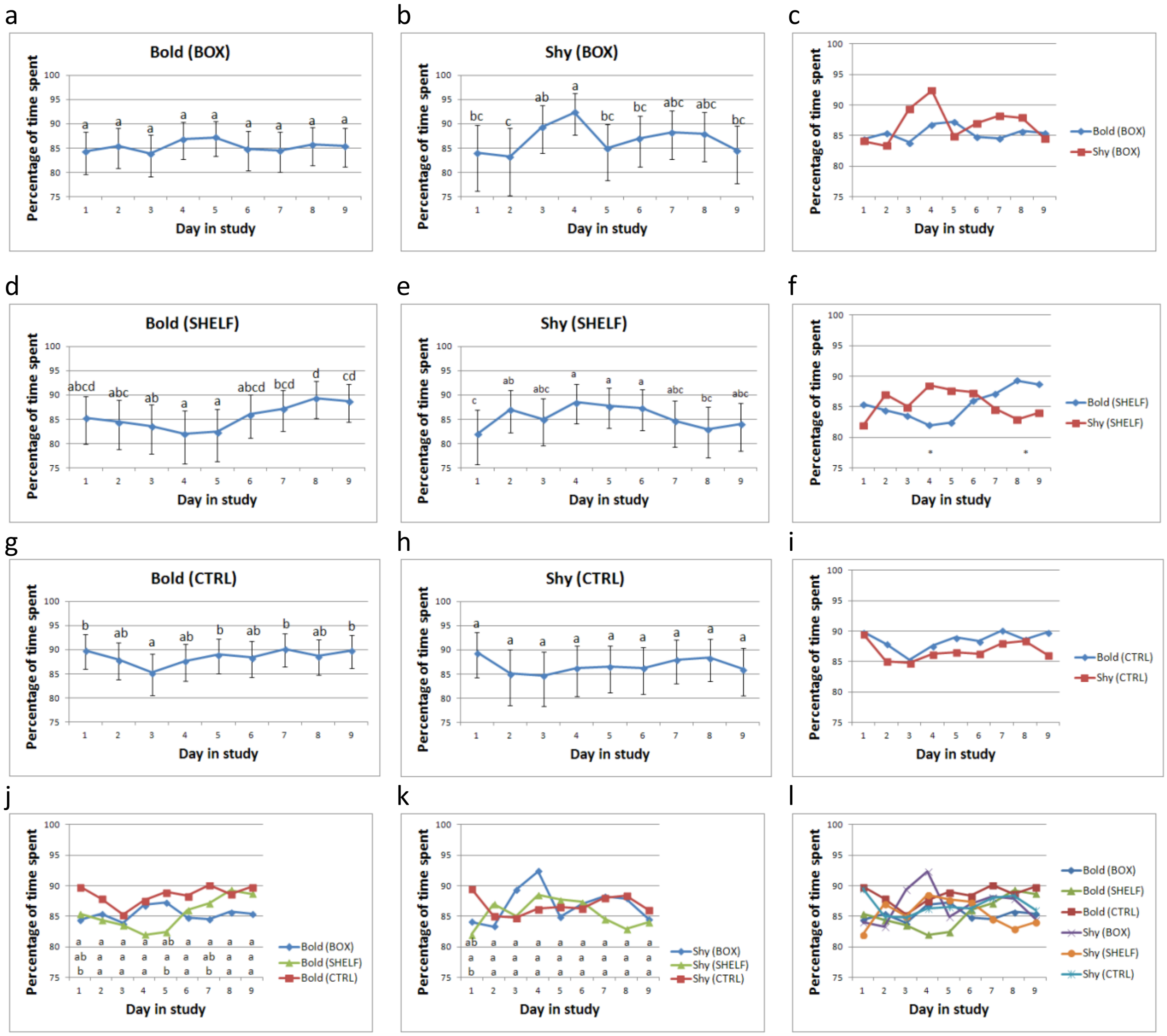

Figure B.5 Combined effects of 'day in study', 'mode of behavioural style' and 'treatment group' on percentage of time spent lying. Results presented are of back-transformed means, and the error bars are 95% confidence intervals. Figures B.5a-c present BOX cats, Figures B.5d-f present SHELF cats, Figures B.5g-i present CONTROL cats, and Figures B.5j-l present all treatment groups together. The first column in each row shows the changes in bold cats over time, and the second column shows the changes in shy cats over time; treatment groups with the same letter code were not significantly different (P>0.05). The final column shows the differences between bold and shy cats over time. These figures uses asterisks to indicate significant difference between bold and shy cats on particular days (P<0.05) (no significant differences were found).